



**PERBANDINGAN KINERJA METODE *MACHINE LEARNING* ANTARA
MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* TERHADAP
PERSEPSI PENGGUNA JASA TRANSPORTASI *ONLINE***

TUGAS AKHIR

Muhamad Dafa Adzikri
41518010174

UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022



**PERBANDINGAN KINERJA METODE *MACHINE LEARNING* ANTARA
MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* TERHADAP
PERSEPSI PENGGUNA JASA TRANSPORTASI *ONLINE***

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Muhamad Dafa Adzikri

41518010174

UNIVERSITAS
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

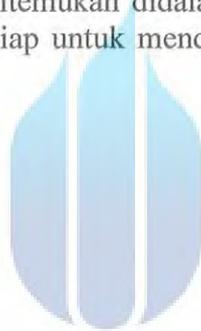
Yang bertanda tangan dibawah ini:

NIM : 41518010174

Nama : Muhamad Dafa Adzikri

Judul Tugas Akhir : PERBANDINGAN KINERJA METODE *MACHINE LEARNING* ANTARA MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* TERHADAP PERSEPSI PENGGUNA JASA TRANSPORTASI *ONLINE*

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



UNIVERSITAS
MERCU BUANA

Jakarta, 29 Juni 2022



Muhamad Dafa Adzikri

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Muhamad Dafa Adzikri
NIM : 41518010174
Judul Tugas Akhir : **PERBANDINGAN KINERJA METODE
MACHINE LEARNING ANTARA MODEL
NAIVE BAYES DAN K-NEAREST NEIGHBOR
TERHADAP PERSEPSI PENGGUNA JASA
TRANSPORTASI ONLINE**

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/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.

arta, 29 Juni 2022

<Muhamad Dafa Adzikri>

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Muhamad Dafa Adzikri
 NIM : 41518010174
 Judul Tugas Akhir : PERBANDINGAN KINERJA METODE
MACHINE LEARNING ANTARA MODEL *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR*
 TERHADAP PERSEPSI PENGGUNA JASA
 TRANSPORTASI *ONLINE*

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			
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika : Jurnal Pengembangan IT			
	ISSN	: 2477-5126 (print), 2548-9356 (online)			
	Link Jurnal	: ejournal.politektegal.ac.id			
	Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. 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, 22 Juni 2022



<Muhamad Dafa Adzikri>

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010174
Nama : Muhamad Dafa Adzikri
Judul Tugas Akhir : PERBANDINGAN KINERJA METODE
MACHINE LEARNING ANTARA MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR*
TERHADAP PERSEPSI PENGGUNA JASA
TRANSPORTASI *ONLINE*

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, 26 Juli 2022



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

MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010174
Nama : Muhamad Dafa Adzikri
Judul Tugas Akhir : PERBANDINGAN KINERJA METODE
MACHINE LEARNING ANTARA MODEL *NAIVE*
BAYES DAN *K-NEAREST NEIGHBOR*
TERHADAP PERSEPSI PENGGUNA JASA
TRANSPORTASI ONLINE

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, 26 Juli 2022



(Wawan Gunawan, S.Kom., MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010174
Nama : Muhamad Dafa Adzikri
Judul Tugas Akhir : PERBANDINGAN KINERJA METODE
MACHINE LEARNING ANTARA MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR*
TERHADAP PERSEPSI PENGGUNA JASA
TRANSPORTASI *ONLINE*

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, 26 Juli 2022



(Dwi Anindyani Rocmah, ST, MTI)

MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518010174
Nama : Muhamad Dafa Adzikri
Judul Tugas Akhir : PERBANDINGAN KINERJA METODE MACHINE LEARNING ANTARA MODEL NAIVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP PERSEPSI PENGGUNA JASA TRANSPORTASI ONLINE

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, 26 Juli 2022

Menyetujui,



(Elyani Dr. Ir.)
Dosen Pembimbing

Mengetahui,



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



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

UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Perbandingan Kinerja Metode *Machine Learning* Antara Model *Naïve Bayes* dan *K-Nearest Neighbor*” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua. Ayah dan Ibu, yang tak pernah lelah mendukung dan selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik, juga tak pernah luput mendoakan yang terbaik untuk proses meraih gelar sarjana bagi saya.
2. Ibu Eliyani, Dr. Ir. selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai
3. Bapak Muhammad Rifqi, S.Kom, M.Kom selaku Dosen Pembimbing Akademik
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Sahabat dan kerabat, yang telah percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana dengan baik, juga tidak pernah bosan memberi dukungan dan doa.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
7. Rana Sahara yang tidak pernah lelah membantu kapanpun dan dimanapun.
8. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR....	ii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PERSETUJUAN PENGUJI	vi
LEMBAR PERSETUJUAN PENGUJI	vii
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA.....	10
BAB 1. LITERATUR REVIEW	11
BAB 2. ANALISIS DAN PERANCANGAN.....	22
BAB 3. SOURCE CODE	27
BAB 4. DATASET.....	33
BAB 5. TAHAPAN EKSPERIMEN.....	35
BAB 6. HASIL SEMUA EKSPERIMEN.....	42
DAFTAR PUSTAKA	68
LAMPIRAN DOKUMEN HAKI.....	70
LAMPIRAN KORESPONDENSI	72

PERBANDINGAN KINERJA METODE *MACHINE LEARNING* ANTARA MODEL *NAIVE BAYES* DAN *K-NEAREST NEIGHBOR* TERHADAP SENTIMEN PENGGUNA JASA TRANSPORTASI *ONLINE*

Muhamad Dafa Adzikri^{1*)}, Eliyani²

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercubuana, Jakarta

^{1,2}Jln. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Kota Jakarta, 11650, Indonesia

email: ¹dafa.adzikri4@gmail.com, ²eliyani@mercubuana.ac.id

Abstract – Technological developments have progressed rapidly, especially in the field of transportation. Unlike the previous habit, nowadays people are more assisted and facilitated by online transportation service technology in supporting daily mobility by using a smartphone. On the other hand, there are shortcomings in the use of online transportation service technology, such as a rate that increases, crime, and the lack of consumer privacy rights, this has an impact on the number of opinions or 'tweets' from the Twitter user community, so a system is needed to analyze the public perception of online transportation service users. This study aims to compare the algorithms between Naïve Bayes and K-Nearest Neighbor in classifying Twitter sentiment datasets and compare manual and automatic label data processes, and compare the percentage of data separation between 10%, 20%, 30% on the test data. In this study, there were five main stages carried out, including the process of data collection, data labeling, text preprocessing, training process and model testing. Based on the results of the study that using manual labeling is more representative in determining the class of data. The 10% data separation percentage model has the highest accuracy value, so in this study it can be seen that the more data that becomes training data, the accuracy of the algorithm increases for the better. The naive Bayes algorithm and k-nearest neighbor algorithm can classify the data well, but the naive Bayes algorithm gets the highest accuracy value of 95% and the k-nearest neighbor algorithm gets the highest accuracy of 93%.

Abstrak – Perkembangan teknologi mengalami kemajuan pesat, khususnya pada bidang transportasi. Berbeda dengan kebiasaan sebelumnya, saat ini masyarakat lebih terbantu dan termudahkan oleh teknologi jasa transportasi online dalam menunjang mobilitas sehari-hari hanya dengan menggunakan smartphone. Disisi lain, terdapat kekurangan dalam penggunaan teknologi jasa transportasi online, seperti halnya kenaikan tarif, kriminalitas, dan kurangnya hak privasi konsumen, hal ini berdampak banyaknya pendapat atau 'cuitan' dari masyarakat pengguna Twitter, sehingga

dibutuhkan sistem untuk menganalisis persepsi masyarakat pengguna jasa transportasi online. Penelitian ini bertujuan untuk membandingkan algoritma antara Naïve Bayes dan K-Nearest Neighbor dalam mengklasifikasi dataset sentimen Twitter serta membandingkan proses data label manual dan otomatis, dan membandingkan presentase pemisahan data antara 10%, 20%, 30% pada data uji. Dalam penelitian ini terdapat lima tahapan utama yang dilakukan, diantaranya mencakup proses pengumpulan data, pelabelan data, text preprocessing, proses pelatihan dan pengujian model. Berdasarkan hasil dari penelitian bahwa menggunakan labeling manual mendapat nilai akurasi yang lebih tinggi dari labeling otomatis. Model presentase pemisahan data 10% memiliki nilai akurasi tertinggi. Pada metode naive bayes mendapatkan nilai akurasi tertinggi sebesar 95% dan algoritma k-nearest neighbor mendapat akurasi tertinggi sebesar 93%.

Kata Kunci – Transportasi Online, Naïve Bayes, K-Nearest Neighbor, Machine Learning.

PENDAHULUAN

Perkembangan teknologi telah mengalami kemajuan yang sangat pesat khususnya pada *electronic commerce (e-commerce)*. Banyak sekali kenyamanan yang diberikan oleh *platform e-commerce* yang membuat banyak pihak meninggalkan transaksi dengan tradisi lama, yaitu dengan tatap muka. Platform yang sering digunakan adalah sebuah platform yang memberikan jasa transportasi. Berbeda dari jasa transportasi pada umumnya, jasa transportasi *online* ini memanfaatkan sistem *online* untuk menghubungkan *driver* dengan konsumen.

Aplikasi jasa transportasi *online* ini memiliki beragam pelayanan, diantaranya adalah pemesanan jasa antar jemput, makanan, dan barang belanja yang sangat membantu meminimalisir waktu. Kemudahan didalamnya

menggunakan fitur yang beragam dan juga banyak terdapat pada *smartphone* yang membuat jasa transportasi *online* banyak diminati

Berdasarkan survei pada tahun 2018 oleh Yayasan Lembaga Konsumen Indonesia (YLKI), 41% konsumen transportasi online pernah dikecewakan, bahkan kini terbukti transportasi online tidak nyaman dan seaman yang dibayangkan sebelumnya[1]. Mulai dari kenaikan tarif dan berbagai kriminalitas, seperti pembunuhan dan perampasan banyak terjadi pada jasa transportasi online dan korbannya adalah konsumen. Terdapat beberapa kelemahan pada transportasi online yaitu tarif transportasi online yang mahal, kurangnya hak privasi dari konsumen, seperti *driver* yang suka menyimpan lalu menghubungi nomor konsumen melalui *Whatsapp* yang berakibat banyaknya ‘cuitan’ dari masyarakat pengguna aplikasi *Twitter* yang berupa kritikan serta keluhan terhadap beberapa perusahaan jasa transportasi *online*. Masalah di atas mengenai transportasi *online* memiliki kesan negatif oleh masyarakat, sehingga hal ini menarik untuk diteliti maka dibuatlah sebuah sistem untuk menganalisis persepsi mengenai pengguna jasa transportasi online.

Analisis penelitian ini menggunakan model Machine Learning *Naives Bayes* dan *K-Nearest Neighbor* yang mengklasifikasikan persepsi masyarakat terhadap penggunaan jasa transportasi *online*. Data yang digunakan didapat dari *tweet* masyarakat yang berisikan tentang persepsi masyarakat terhadap pengguna jasa transportasi *online*. Tujuan dari analisis ini ialah untuk mengumpulkan polaritas melalui teks atau opini pada dataset yang bersifat positif, negatif dan netral dengan menggunakan dua data label dengan membandingkan data label manual dan label otomatis serta membandingkan *split validation* yang terbaik untuk mengetahui model mana yang lebih baik dalam tingkat keakurasian yang dihasilkan.

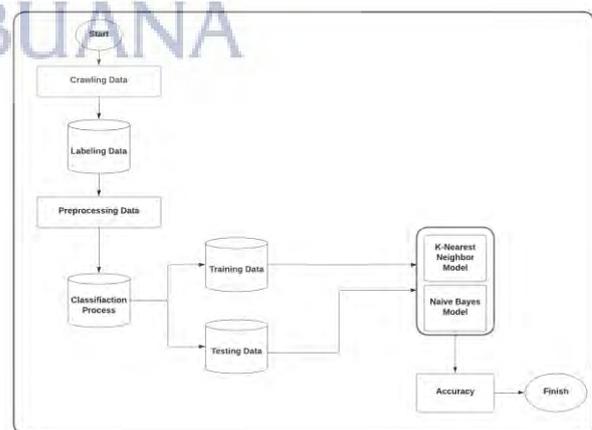
PENELITIAN YANG TERKAIT

Penelitian mengenai perbandingan algoritma antara *decision tree (DT)* dan *support vector machine (SVM)* telah dilakukan oleh [2] dalam analisis sentimen dimana hasil dari penelitian ini ditemukan nilai *k-fold* terbaik untuk model *support vector machine* dan *decision tree* dimana kinerja algoritma *support vector machine* mendapat nilai akurasi yang lebih baik sebesar 91,67% dibandingkan algoritma *decision tree* sebesar 89,20%. Selanjutnya, dilakukan penelitian mengenai fitur seleksi *particle swarm optimization* oleh [3] hasil yang didapat bahwa fitur seleksi *particle swarm optimization* dapat meningkatkan hasil akurasi dari algoritma *support vector machine* dimana nilai akurasi semula sebesar 87% setelah menggunakan fitur seleksi *particle swarm optimization* akurasi meningkat menjadi 98%. Penelitian selanjutnya dilakukan oleh [4] pada penelitian ini dapat diketahui kata paling banyak muncul seperti kata “ovo” kata terbanyak muncul pada sentiment positif, dan kata “driver” pada kata negatif yang

menandakan keluhan terhadap *driver Grab*, sehingga dapat menjadi evaluasi bagi perusahaan *Grab*. Lalu, penelitian selanjutnya dilakukan oleh [5] Pada penelitian ini ditemukan kata-kata yang sering muncul pada sentimen negatif yang berhubungan mengenai lokasi peta dan notifikasi akun terdapat error. Sehingga dapat menjadi bahan evaluasi bagi perusahaan *Gojek*. Selanjutnya, dilakukan penelitian mengenai analisis sentimen dengan implementasi dari teknik fitur seleksi *particle swarm optimization* oleh [6] dimana hasil pengujian yang dilakukan penelitian ini dapat disimpulkan bahwa *particle swarm optimization (PSO)* dapat meningkatkan *performance* dari algoritma *naive bayes* sebesar 2,84% dari 69,50% menjadi 72,34% nilai akurasi yang didapat. Penelitian selanjutnya dilakukan oleh [7] dalam klasifikasi analisis sentimen menggunakan algoritma *k-nearest neighbor* dimana algoritma *k-nearest neighbor* mendapat nilai akurasi terbaik pada nilai $k=5$, memiliki nilai akurasi sebesar 94,12%, dengan error sebesar 5,88%. Penelitian selanjutnya, dilakukan oleh [8] pada penelitian ini dilakukan analisis sentimen pada aplikasi tiket *online* menggunakan algoritma *support vector machine*, penelitian ini menganalisis sentimen komentar pada *Google Playstore* dimana aplikasinya adalah Pegipegi, Agoda, Traveloka. Dimana sentimen pada aplikasi Pegipegi nilai akurasi yang didapat sebesar 78,21%, pada aplikasi Agoda sebesar 77% dan aplikasi Traveloka mendapat nilai akurasi sebesar 75,03%.

METODE PENELITIAN

Dalam penelitian ini terdapat lima tahap utama yang dilakukan, diantaranya mencakup proses pengumpulan data, *labeling data*, *text preprocessing*, proses pelatihan dan pengujian model. Proses klasifikasi menggunakan dua metode, yaitu *naive bayes* dan *k-nearest neighbor*. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1 Tahapan Penelitian

A. Crawling Data

Crawling data dalam penelitian ini bertujuan untuk mengumpulkan data dari *server twitter*. Pengumpulan data

dilakukan dengan memanfaatkan fasilitas *application programming interface (API)* yang di sediakan oleh *twitter*. Kata kunci yang digunakan dalam proses *crawling* adalah “Transportasi Online”, “Gojek”, “Grab”, “Ojol”, “Shopeefood”, “Gofood”, “Grabfood”, “Gosend”, “Goto”. Data yang didapat sebanyak 5000 data diperoleh kemudian disimpan dalam bentuk *.csv tweet* data diambil dari tanggal 1 april 2022 sampai 31 mei 2022 berkaitan dengan respon masyarakat dimasa endemi mengenai transportasi online. Berikut adalah contoh data yang didapat dari proses *crawling data* pada tabel 1.

TABEL I
CONTOH *CRAWLING DATA*.

Nomor	Tweet
1	RT @babygrace pake gojek aja ok nanti kamu marah kalo telat
2	haduh macet banget tadi pagi gegara banyak ojol, bikin penuh aja #macet
3	RT Cek @key pake aplikasi hijau aja gojek atau grab cok
4	Pake ojol keknya enak hari ini #gogojek #santai
5	Enaknya ada gojek pagi2 bisa nyantai ga perlu nyeti #gogojek #ojol

B. Labeling Data

Labeling data dilakukan untuk menentukan kelas positif, negatif, dan netral. Pada penelitian ini dilakukan dua tahapan *labeling* yang pertama dilakukan *manual* oleh peneliti, dan kedua menggunakan *library textblob* dengan menilai *polarity* dan *subjectivity* suatu teks untuk mengetahui kelas positif, negatif, dan netral. Berikut adalah tabel hasil dari label data

TABEL II
HASIL *LABELING DATA*

Tweet	Kelas
Gila ya rame banget ojol bikin macet aja njir	Negatif
Bener kan kalo pulang lebih enak naik gojek daripada naik busway bisa pulang pas ada orang di jalan	Positif
Ngaburide sama abang gojek hujan-hujan	Netral

C. Preprocessing Data

Tujuan dari *preprocessing data* adalah untuk membersihkan data, integrasi data, transformasi data dan reduksi data[9]. Adapun beberapa tahapan *preprocessing* yang dilakukan seperti *data cleansing*, *tokenization*, *text normalization*, *case folding*, *stopword removal*, dan *stemming*[10].

1. *Data Cleansing*
Data Cleansing bertujuan untuk menghapus RT, *hashtag*, angka, simbol dan URL pada tweets[11].
2. *Case Folding*
Case Folding bertujuan untuk merubah teks dari penulisan huruf besar menjadi penulisan huruf kecil[12].
3. *Tokenizing*
Tokenizing dilakukan untuk memisahkan kata untuk memudahkan proses analisis[13].
4. *Text Normalization*
Teks Normalization bertujuan untuk mengubah kata-kata “gaul” menjadi kalimat yang tepat[14].
5. *Stopword Removal*
Stopword Removal bertujuan untuk menghapus kata penghubung dalam penelitian ini peneliti menggunakan *stopword* yang berasal dari *corpus* sastra[15].
6. *Stemming*
Stemming merupakan tahapan terakhir dalam *preprocessing* dalam penelitian ini yang berfungsi untuk mengubah kata menjadi kata dasar[16].

Berdasarkan tahap-tahap yang telah dijabarkan. Berikut adalah contoh hasil *output* dari setiap tahapan yang dilakukan terdapat pada tabel berikut:

Tabel III
Contoh Hasil Dari Tahapan *Preprocessing*

Tahapan	Teks
Teks Awal	Aku naik Gojek atau TJ aja ya yg enak males kalo kemacetan #magerbanget
<i>Data Cleansing</i>	Aku naik Gojek atau TJ aja ya yg enak males kalo kemacetan magerbanget
<i>Case Folding</i>	aku naik gojek atau tj aja ya yg enak males kalo kemacetan magerbanget
<i>Tokenizing</i>	[aku, naik, gojek, atau, tj, aja, ya, yg, enak, males, kalo, kemacetan, mager, banget]
<i>Text Normalization</i>	[aku, naik, gojek, atau, tj, aja, ya, yang, enak, males, kalo, kemacetan, mager, banget]
<i>Stopword Removal</i>	[naik, gojek, tj, enak, males, kemacetan, mager, banget]
<i>Stemming</i>	[naik, gojek, tj, enak, males, macet, mager, banget]

D. Proses Klasifikasi Data

Pada proses klasifikasi data, dilakukan *split data* menjadi *data test* sebagai data uji dan *data train* sebagai

data latih. Pada penelitian ini dilakukan tiga percobaan pemisahan data untuk mengetahui pemisahan data yang terbaik dalam memaksimalkan nilai akurasi, yakni memisahkan sebanyak 10%, 20% dan 30% pemisahan data.

E. Klasifikasi Data Menggunakan Algoritma *Naive Bayes*

Naive bayes merupakan metode klasifikasi yang dapat memprediksi suatu kelas sehingga dapat menghasilkan keputusan berdasarkan data latih atau pembelajaran[17]. *Naive bayes* memiliki kelebihan, diantaranya sederhana, menghasilkan akurasi yang tinggi ketika di terapkan pada data yang besar dan juga cepat[18]. Secara umum, klasifikasi *naive bayes* dapat dilihat pada persamaan 1.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

F. Klasifikasi Data Menggunakan Algoritma *K-Nearest Neighbor*

K-nearest neighbor merupakan salah satu algoritma yang terdapat pada *machine learning*. Algoritma ini memiliki tujuan untuk mengklasifikasikan suatu *dataset* pada kelas yang telah ditetapkan dari kelompok sampel yang telah dibuat[19]. Klasifikasi *k-nearest neighbor* dapat dilihat pada persamaan 2.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (ar(x_i) - ar(x_j))^2} \quad (2)$$

G. *Confusion Matrix*

Evaluasi akhir penelitian ini menggunakan metode *confusion matrix*. Cara kerja metode ini yaitu matriks dari prediksi akan dibandingkan dengan kelas asli yang berisi informasi nyata dan prediksi nilai klasifikasi[20]. Model *confusion matrix* dapat dilihat pada berikut:

Tabel IV
Model *Confusion Matrix*

Predicted Value	Actual Value		
	0(Negatif)	1(Neutral)	2(Positif)
0(Negatif)	TP(True Positive)	FP(False Positive)	FP(False Positive)
1(Neutral)	FP(False Positive)	TP(True Positive)	FP(False Positive)
2(Positif)	FP(False Positive)	FP(False Positive)	TP(True Positive)

Hasil akhir yang didapat menjadi nilai akurasi dapat dilihat melalui persamaan 3.

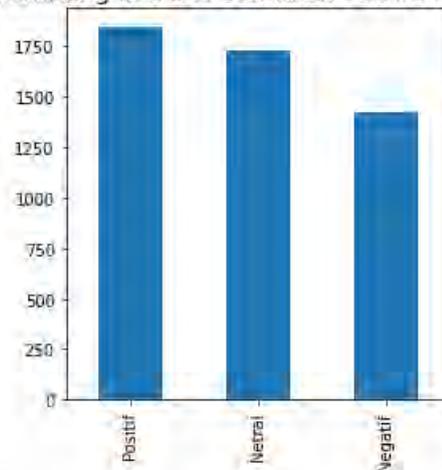
$$Accuracy = \frac{TP1+TP2+TP3}{FP1+FP2+FP3+FP4+FP5+FP6+TP1+TP2+TP3} \quad (3)$$

HASIL DAN PEMBAHASAN

Berdasarkan metode penelitian yang telah dirancang. Tahapan yang dilakukan pertama membandingkan hasil

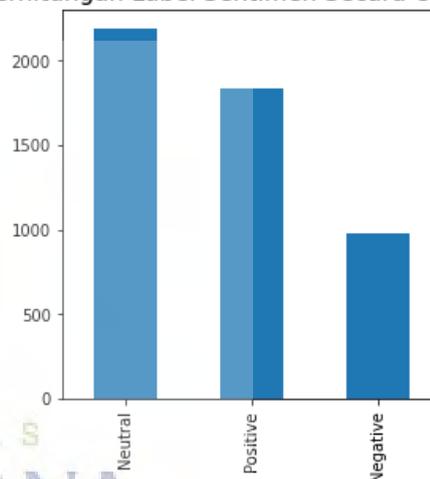
label manual dengan label otomatis dapat dilihat pada gambar 2 dan 3.

Gambar 2 Grafik Perhitungan Label Secara Manual



Gambar 2 Grafik Perhitungan Label Secara Manual

Gambar 3 Grafik Perhitungan Label Secara Otomatis



Gambar 3 Grafik Perhitungan Label Secara Otomatis

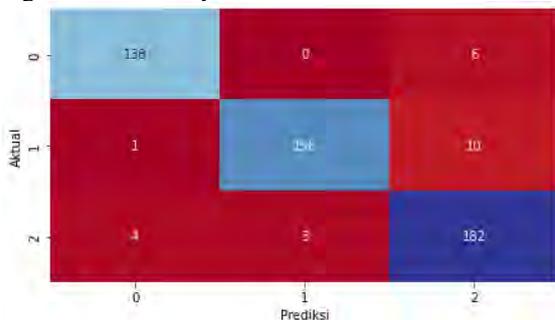
Berdasarkan data yang dipaparkan dapat dilihat perbedaan yang cukup jelas antara data label manual dan data label otomatis, dimana pada label manual kelas data menunjukkan hasil terbanyak adalah kelas data positif namun pada label secara otomatis kelas data menunjukkan hasil terbanyak adalah kelas data netral, sehingga peneliti melakukan percobaan dengan membandingkan dua label *dataset* untuk hasil klasifikasi terbaik

A. Implementasi Algoritma *Naive Bayes*

Pada penelitian ini dilakukan perbandingan *percentage* pada pemisahan data sehingga dapat diketahui pemisahan data terbaik untuk mengklasifikasi data

1. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 10% Labeling Manual*

Percobaan pertama yang dilakukan adalah membagi data sebanyak 10% dimana data uji sebanyak 10% dan data latih sebanyak 90%. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes*.



Gambar 4. Performa Algoritma *Naïve Bayes* Pada Data Label Manual 10%

Melalui data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

$$Accuracy = \frac{Tp1+Tp2+Tp3}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{138+156+182}{500} = \frac{476}{500} = 95\% \quad (4)$$

Hasil perhitungan yang terdapat pada persamaan 4, didapatkan hasil akurasi sebesar 95%.

2. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 10% Labeling Otomatis*

Percobaan kedua yang dilakukan adalah membagi data sebanyak 10% pada data label otomatis. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes* pada data label otomatis.



Gambar 5. Performa Algoritma *Naïve Bayes* Pada Data Label Otomatis 10%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

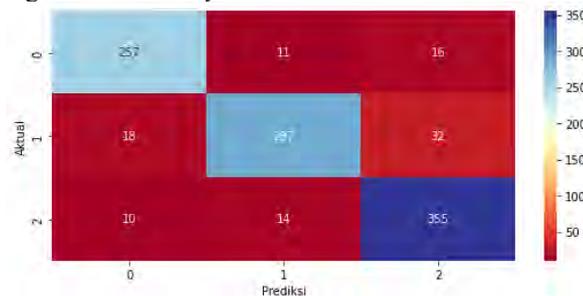
$$Accuracy = \frac{Tp1+Tp2+Tp3}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{45+175+167}{500} = \frac{425}{500} = 85\% \quad (5)$$

Melalui perhitungan yang terdapat pada persamaan 5, didapatkan hasil akurasi sebesar 85%.

3. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 20% Labeling Manual*

Percobaan ketiga yang dilakukan adalah membagi data sebanyak 20% dimana data uji sebanyak 20% dan data latih sebanyak 80%. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes*



Gambar 6. Performa Algoritma *Naïve Bayes* Pada Data Label Manual 20%

Melalui data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

$$Accuracy = \frac{Tp1+Tp2+Tp3}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{257+287+355}{1000} = \frac{900}{1000} = 90\% \quad (6)$$

Hasil perhitungan yang terdapat pada persamaan 6, didapatkan hasil akurasi sebesar 90%.

4. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 20% Labeling Otomatis*

Percobaan keempat yang dilakukan adalah membagi data sebanyak 20% pada data label otomatis. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes* pada data label otomatis.



Gambar 7. Performa Algoritma *Naïve Bayes* Pada Data Label Otomatis 20%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

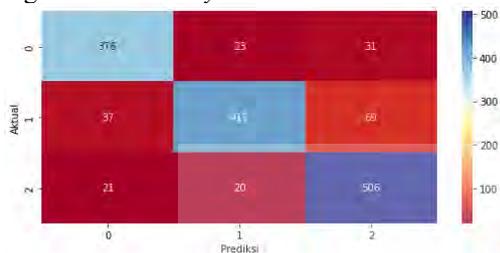
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{102+419+302}{1000} = \frac{820}{1000} = 82\% \quad (7)$$

Melalui perhitungan yang terdapat pada persamaan 7, didapatkan hasil akurasi sebesar 82%.

5. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 30% Labeling Manual*

Percobaan kelima yang dilakukan adalah membagi data sebanyak 30% dimana data uji sebanyak 30% dan data latih sebanyak 70%. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes*



Gambar 8. Performa Algoritma *Naïve Bayes* Pada Data Label Manual 30%

Melalui data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

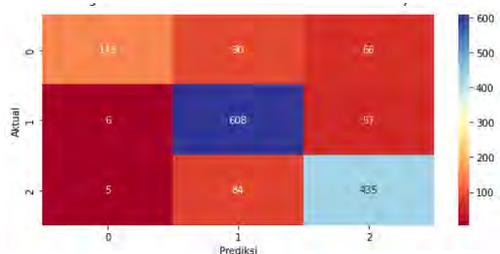
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{376+417+506}{1500} = \frac{1299}{1500} = 87\% \quad (8)$$

Hasil perhitungan yang terdapat pada persamaan 8, didapatkan hasil akurasi sebesar 87%.

6. Klasifikasi Algoritma *Naïve Bayes* Pada *Split Data 30% Labeling Otomatis*

Percobaan keenam yang dilakukan adalah membagi data sebanyak 30% pada data label otomatis. Berikut adalah performa hasil klasifikasi dari algoritma *naïve bayes* pada data label otomatis.



Gambar 9. Performa Algoritma *Naïve Bayes* Pada Data Label Otomatis 30%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

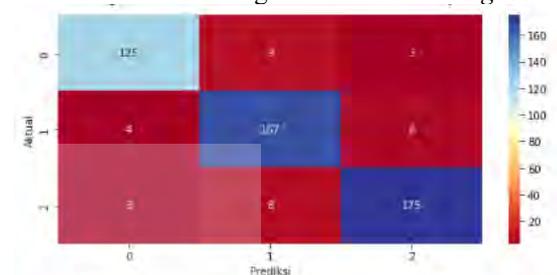
$$Accuracy = \frac{149+608+435}{1500} = \frac{1192}{1500} = 79\% \quad (9)$$

Melalui perhitungan yang terdapat pada persamaan 9 didapatkan hasil akurasi sebesar 79%.

B. Implementasi Algoritma *K-Nearest Neighbor*

1. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 10% Labeling Manual*

Percobaan pertama pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 10% dimana data uji sebanyak 10% dan data latih sebanyak 90%. Berikut adalah performa hasil klasifikasi dari algoritma *k-nearest neighbor*



Gambar 10. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Manual 10%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

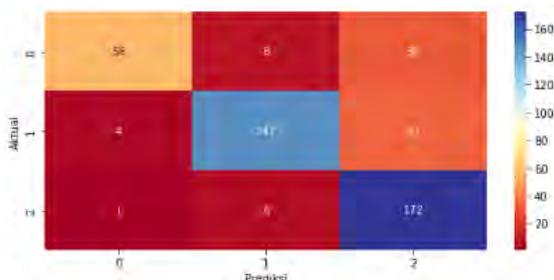
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{125+167+175}{500} = \frac{467}{500} = 93\% \quad (10)$$

Melalui hasil perhitungan yang terdapat pada persamaan 10, didapatkan hasil akurasi sebesar 93%.

2. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 10% Labeling Otomatis*

Percobaan kedua pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 10% pada data label otomatis. Berikut adalah performa hasil klasifikasi pada data label otomatis



Gambar 11. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Otomatis 10%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

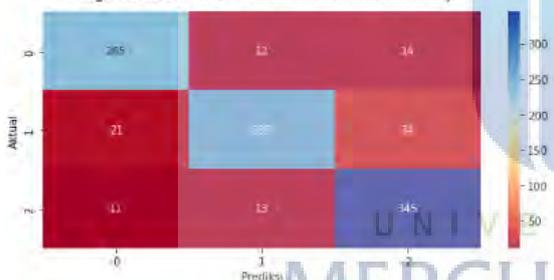
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{56+147+172}{500} = \frac{410}{500} = 82\% \quad (11)$$

Melalui perhitungan yang terdapat pada persamaan 11, didapatkan hasil akurasi sebesar 82%.

3. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 20% Labeling Manual*

Percobaan ketiga pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 20% dimana data uji sebanyak 20% dan data latih sebanyak 80%. Berikut adalah performa hasil klasifikasi dari algoritma *k-nearest neighbor*



Gambar 12. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Manual 20%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

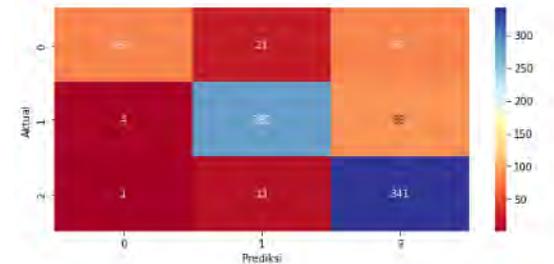
$$Accuracy = \frac{265+285+345}{1000} = \frac{903}{1000} = 90\% \quad (12)$$

Melalui hasil perhitungan yang terdapat pada persamaan 12, didapatkan hasil akurasi sebesar 90%.

4. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 20% Labeling Otomatis*

Percobaan keempat pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 20% pada data label otomatis. Berikut

adalah performa hasil klasifikasi pada data label otomatis



Gambar 13. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Otomatis 20%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

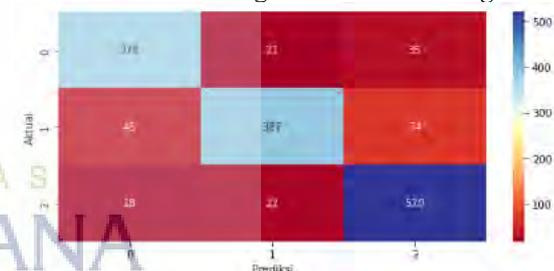
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{81+282+341}{1000} = \frac{771}{1000} = 77\% \quad (13)$$

Melalui perhitungan yang terdapat pada persamaan 13, didapatkan hasil akurasi sebesar 77%.

5. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 30% Labeling Manual*

Percobaan kelima pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 30% dimana data uji sebanyak 30% dan data latih sebanyak 70%. Berikut adalah performa hasil klasifikasi dari algoritma *k-nearest neighbor*



Gambar 14. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Manual 30%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

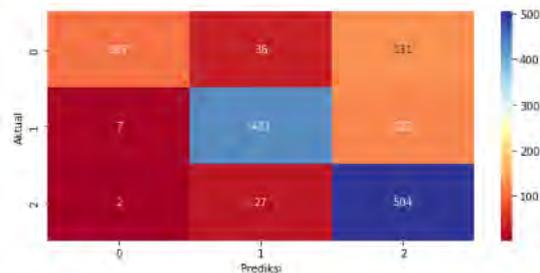
$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{378+387+520}{1500} = \frac{1285}{1500} = 86\% \quad (14)$$

Melalui perhitungan yang terdapat pada persamaan 14, didapatkan hasil akurasi sebesar 86%.

6. Klasifikasi Algoritma *K-Nearest Neighbor* Pada *Split Data 30% Labeling Otomatis*

Percobaan keempat pada algoritma *k-nearest neighbor* yang dilakukan adalah membagi data sebanyak 30% pada data label otomatis. Berikut adalah performa hasil klasifikasi pada data label otomatis



Gambar 15. Performa Algoritma *K-Nearest Neighbor* Pada Data Label Otomatis 30%

Berdasarkan data yang dipaparkan nilai akurasi dapat dihitung melalui persamaan berikut:

$$Accuracy = \frac{Tp}{Jumlah\ Semua\ Data}$$

$$Accuracy = \frac{107+433+504}{1500} = \frac{1044}{1500} = 76\% \quad (15)$$

Melalui perhitungan yang terdapat pada persamaan 15, didapatkan hasil akurasi sebesar 76%.

C. Komparasi Analisis Hasil Prediksi

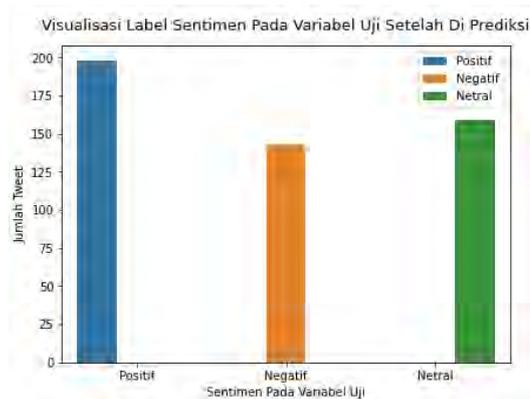
Berdasarkan pengujian yang telah dilakukan dengan hasil akhir berupa nilai prediksi dari dua model pengujian terhadap dua data label yang berbeda dapat dilihat melalui tabel 5.

Tabel V
Perbandingan Hasil Akhir Data

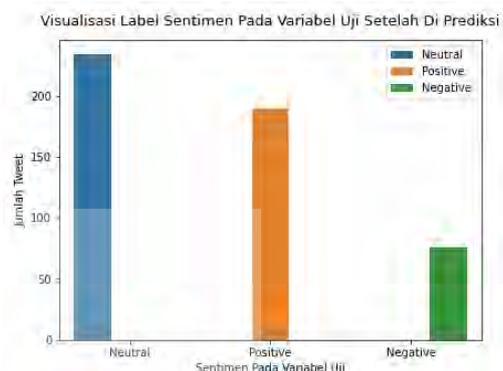
Percobaan	Label Manual		Label Otomatis	
	KNN	NB	KNN	NB
10%	93%	95%	82%	85%
20%	90%	90%	77%	82%
30%	86%	87%	76%	79%

D. Hasil Prediksi Sentimen

Berdasarkan pengujian yang telah dilakukan prediksi akhir berupa grafik dari dua label data yang berbeda dapat dilihat melalui gambar 15 dan 16.



Gambar 15. Visualisasi Prediksi Pada Variabel Uji Label Manual



Gambar 16. Visualisasi Prediksi Pada Variabel Uji Label Otomatis

KESIMPULAN

Pada penelitian ini kesimpulan yang didapatkan yaitu:

1. Berdasarkan hasil dari penelitian dapat dilihat bahwa *labeling* manual mendapat nilai akurasi lebih tinggi dari *labeling* otomatis
2. Split validation 10% memiliki tingkat akurasi tertinggi
3. Algoritma *naïve bayes* dan *k-nearest neighbor* dapat mengklasifikasikan data dengan baik, namun algoritma *naïve bayes* mendapatkan nilai akurasi tertinggi sebesar 95% dan algoritma *k-nearest neighbor* mendapat akurasi tertinggi sebesar 93%.
4. Hasil prediksi dari label manual menunjukkan sentimen positif
5. Hasil prediksi dari label Otomatis menunjukkan sentimen netral

DAFTAR PUSTAKA

- [1] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro*

- dan *Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [2] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
- [3] B. Rahmatullah, P. Budiyono, and S. A. Saputra, “Sentimen Analisis Transportasi Online Menggunakan Algoritma Support Vector Machine, Naive Bayes, dan KNN,” *J. Ilmu Komput. JIK*, vol. IV, no. 02, pp. 9–15, 2021.
- [4] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [5] M. N. Muttaqin and I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021.
- [6] I. Wahyudi, S. Bahri, and P. Handayani, “Aplikasi Pembelajaran Pengenalan Budaya Indonesia,” vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.
- [7] S. Diansyah, “Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN),” *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 7–12, 2022, doi: 10.37034/jsisfotek.v4i1.114.
- [8] F. Bei and S. Sudin, “Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm),” *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [9] A. Budiman, J. C. Young, and A. Suryadibrata, “Implementasi Algoritma Naïve Bayes untuk Klasifikasi Konten Twitter dengan Indikasi Depresi,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 2, pp. 133–138, 2021.
- [10] V. Ayumi and I. Nurhaida, “Klasifikasi Chest X-Ray Images Berdasarkan Kriteria Gejala Covid-19 Menggunakan Convolutional Neural Network,” *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, pp. 147–153, 2021, doi: 10.36085/jsai.v4i2.1513.
- [11] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [12] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [13] B. Laurensz and Eko Sedyono, “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [14] M. F. Putra, A. Herdiani, D. Puspendari, F. Informatika, and U. Telkom, “Analisis Pengaruh Normalisasi , TF-IDF , Pemilihan Feature-set Terhadap Klasifikasi Sentimen Menggunakan Maximum Entropy (Studi Kasus : Grab dan Gojek),” vol. 6, no. 2, pp. 8520–8529, 2019.
- [15] E. B. Santoso and A. Nugroho, “Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook,” *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [16] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [17] A. A. Nabhan, B. Rahayudi, and D. E. Ratnawati, “Klasifikasi Tweets Masyarakat yang Membicarakan Layanan GoFood dan GoRide pada GoJek Dimedia Sosial Twitter Selama Masa Kenormalan Baru (New Normal) dengan Metode Naïve Bayes,” vol. 5, no. 7, pp. 3018–3025, 2021.
- [18] D. F. Sari, D. Kurniawati, E. Prayitno, and I. Irfangi, “Sentiment Analysis of Twitter Social Media to Online Transportation in Indonesia Using Naïve Bayes Classifier,” *J. Int. Conf. Proc.*, vol. 2, no. 1, p. 31, 2019.
- [19] C. Rohwinasakti, Savira; Irawan, Budi; Setianingsih, “Analisis Sentimen Pada Produk Layanan Transportasi Online Menggunakan Metode K-Nearest Neighbor,” vol. 7, no. 3, pp. 9312–9321, 2020.
- [20] Styawati., N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.

KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Perbandingan Kinerja Metode *Machine Learning* Antara Model *Naive Bayes* dan *K-Nearest Neighbor*”. Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan: literatur *review*, analisis perancangan, *source code*, *dataset* yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

- Bagian 1: Literature Review menjabarkan mengenai beberapa jurnal yang terkait dengan penelitian.
- Bagian 2: Analisis dan Perancangan menjabarkan analisis masalah dan analisis model.
- Bagian 3: *Source Code* menjabarkan kumpulan kode di setiap prosesnya mulai dari Membaca data, *Preprocessing*, *Labeling*, Implementasi model, sampai pengujian dan evaluasi.
- Bagian 4: *Dataset* melampirkan pengambilan *dataset*.
- Bagian 5: Tahapan eksperimen berisi pengumpulan data, *Preprocessing*, Implementasi model, dan evaluasi model.
- Bagian 6: Hasil implementasi eksperimen secara keseluruhan yang mencakup jawaban dari skenario percobaan pengujian yang sudah disebutkan pada bagian 5. Bagian ini berguna untuk menemukan model Analisa sentimen terbaik yang dilihat dari beberapa pengujian berdasarkan Opini masyarakat yang positif dan negatif. Kemudian, hasil yang didapat digunakan berupa akurasi, presisi dan recal Pendidikan tatap muka di Indonesia untuk bahan evaluasi atas kebijakan yang diterapkan.

UNIVERSITAS
MERCU BUANA