



**PERBANDINGAN ANALISIS SENTIMEN TERHADAP 2 JASA
EKSPEDISI DI INDONESIA DARI MEDIA SOSIAL *TWITTER*
MENGUNAKAN METODE NAIVE BAYES DAN KNN
(K-NEAREST NEIGHBOR)**

TUGAS AKHIR

Nady Febrilianto
41518010046

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



**PERBANDINGAN ANALISIS SENTIMEN TERHADAP 2 JASA EKSPEDISI DI
INDONESIA DARI MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE
NAIVE BAYES DAN KNN
(K-NEAREST NEIGHBOR)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Nady Febrilianto
41518010046

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama Mahasiswa : Nady Febrilianto
NIM : 41518010046
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa
Ekspedisi Di Indonesia Dari Media Sosial *Twitter*
Menggunakan Metode Naive Bayes Dan KNN (K-
Nearest-Neighbor)

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, 27 Juli 2022



Nady Febrilianto

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Nady Febrilianto
NIM : 41518010046
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial *Twitter* Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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.

Jakarta, 27 Juli 2022



UNIVERSITAS
MERCU BUANA
Nady Febrilianto

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Nady Febrilianto
NIM : 41518010046
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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 Pengembang IT			
	ISSN : 2477-5126			
		Link Jurnal : https://ejournal.poltektegal.ac.id/index.php/informatika		
		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.

Mengetahui
Dosen Pembimbing TA

Jakarta, 27 Juli 2022

ca. n.



Leonard Goeiranto, Dr.,ST.M.Se



Nady Febrilianto

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010046
Nama : Nady Febrilianto
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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, 27 Juli 2022



(Eliyani Dr.Ir)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010046
Nama : Nady Febrilianto
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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, 27 Juli 2022



(Muhammad Rifqi S.Kom, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010046
Nama : Nady Febrilianto
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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, 27 Juli 2022



(Anis Cherid, SE, MTI)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518010046
Nama : Nady Febrilianto
Judul Tugas Akhir : Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (K-Nearest-Neighbor)

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, 27 Juli 2022

o.n Menyetujui,



(Leonard Gochimanto, Dr., ST, M.Sc)
Dosen Pembimbing

Mengetahui,



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



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

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Comparison of Sentiment Analysis towards 2 Expedition Services in Indonesia From Twitter Social Media” 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. Bapak Leonard Goeiranto, Dr., ST, M.Sc 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.
2. 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.
3. Bapak Dr. Harwikarya, MT 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. Senior atas berbagai saran yang diberikan dan junior yang dengan semangat mendukung dikala bertemu.

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.

Jakarta, 27 Juli 2022



Nady Febrilianto



DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... ..	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	iv
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	11
BAB 1. LITERATUR REVIEW	12
BAB 2. ANALISIS DAN PERANCANGAN.....	26
BAB 3. SOURCE CODE	29
BAB 4. DATASET.....	43
BAB 5. TAHAPAN EKSPERIMEN.....	45
BAB 6. HASIL SEMUA EKSPERIMEN.....	51
LAMPIRAN DOKUMEN HAKI.....	64
LAMPIRAN KORESPONDENSI	66

PERBANDINGAN ANALISIS SENTIMEN TERHADAP 2 JASA EKSPEDISI DI INDONESIA DARI MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE NAIVE BAYES DAN KNN (K-NEAREST NEIGHBOR)

1st Nady Febrilianto

Ilmu komputer Informatika

Universitas Mercu Buana

Jakarta, Indonesian

41518010046@student.mercubuana.ac.
id

2nd Leonard Goeirmanto

Ilmu komputer Informatika

Universitas Mercu Buana

Jakarta, Indonesian

leonard@mercubuana.ac.id

Abstract— The *Corona Virus (Covid-19)* pandemic has encouraged an increase in *e-commerce* orders. Some transaction mechanisms that are now widely used include *internet online payments, smart cards, radio frequency identification (RFID), mobile payments,* and others. The increase in digital spending occurs because people prefer to buy their needs online, this is in line with the implementation of government policies to work from home or *Work From Home (WFH)* and school from home. The advancement of *e-commerce* in Indonesia has an impact on the explosion of the number of users of shipping services. *E-commerce* in Indonesia plays an important role in shipping service users. Expedition services are able to deliver to the regions, domestically and abroad, to be able to contribute to the delivery process. Courier services also play an important role in supporting the smooth running of the company's business that requires fast and safe delivery services. There are many popular shipping companies in Indonesia, for example first shipping service and second shipping service. This study used Naïve Bayes and KNN (K- Nearest Neighbor) using the first shipping service and second shipping service expedition service sentiment datasets. The percentage split model 70:30 has the best accuracy results on the data tested by first shipping service and second shipping service. the first shipping service case study has an 89% accuracy value for the KNN (K-Nearest Neighbor) algorithm, the second shipping service case study has a 97% accuracy value for the KNN (K-Nearest Neighbor) algorithm on automatic Labeling. The results of this study show that the jasa ekspedisi kedua case study has an accuracy value of 97% for the KNN (K- Nearest Neighbor) algorithm, automatic labeling has a higher accuracy rate with an average accuracy ratio of 97% which has the best average value of this study.

Keywords— *first shipping service and second shipping service, Sentiment analysis, Naïve Bayes, K-Nearest Neighbor, Sentiment Analysis*

Abstrak— Pandemi *Virus Corona (Covid-19)* telah mendorong peningkatan pesanan *e-commerce*. perkembangan *teknologi* dan *system* informasi yang banyak mendukung proses bisnis, maka dari itu dibutuhkan

pengelolaan data transaksi yang semakin cepat pula. Beberapa mekanisme transaksi yang sekarang banyak dipergunakan antara lain adalah *internet online payment, smart card, radio frequency identification (RFID), mobile payment,* dan lain-lain. Peningkatan belanja digital terjadi karena masyarakat lebih memilih untuk membeli kebutuhannya secara online, hal ini sejalan dengan penerapan kebijakan pemerintah untuk bekerja dari rumah atau *Work From Home (WFH)* dan sekolah dari rumah. Semakin majunya *e-commerce* di Indonesia berdampak meledaknya jumlah pengguna jasa pengiriman. *E-commerce* di Indonesia sangat berperan penting dalam pengguna jasa pengiriman. Jasa ekspedisi mampu mengantarkan ke daerah, domestik maupun luar negeri, untuk dapat berkontribusi dalam proses pengiriman. Jasa kurir juga berperan penting dalam menunjang kelancaran bisnis perusahaan yang membutuhkan jasa pengiriman yang cepat dan aman. Ada banyak perusahaan ekspedisi yang populer di Indonesia misalnya jasa ekspedisi pertama dan jasa ekspedisi kedua. Penelitian ini menggunakan Naïve Bayes serta KNN (K- Nearest Neighbor) menggunakan dataset sentimen jasa ekspedisi jasa ekspedisi pertama dan jasa ekspedisi kedua. Model percentage split 70:30 memiliki hasil akurasi terbaik pada Data yang diuji eksperimen jasa ekspedisi pertama dan jasa ekspedisi kedua. studi kasus jasa ekspedisi pertama memiliki nilai akurasi 89% untuk algoritma KNN (K-Nearest Neighbor), studi kasus jasa ekspedisi kedua memiliki nilai akurasi 97% untuk algoritma KNN (K-Nearest Neighbor) pada Labeling otomatis. Pada hasil Penelitian ini menunjukkan studi kasus jasa ekspedisi kedua memiliki nilai akurasi 97% untuk algoritma KNN (K- Nearest Neighbor) Labeling otomatis memiliki hasil tingkat akurasi yang lebih tinggi dengan nilai rata-rata akurasi sebesar 97% yang memiliki nilai rata-rata paling baik dari penelitian ini.

Kata Kunci— *jasa ekspedisi pertama dan jasa ekspedisi kedua, Sentimen, Naïve Bayes, K-Nearest Neighbor, Analisa Sentimen*

I. INTRODUCTION

Pandemi *Virus Corona (Covid-19)* telah mendorong peningkatan pesanan *e-commerce*. perkembangan *teknologi* dan system informasi yang banyak mendukung proses bisnis, maka dari itu dibutuhkan pengelolaan data transaksi yang semakin cepat pula. Beberapa mekanisme transaksi yang sekarang banyak dipergunakan antara lain adalah *internet online payment, smart card, radio frequency identification (RFID), mobile payment*, dan lain-lain[1].

Peningkatan belanja digital terjadi karena masyarakat lebih memilih untuk membeli kebutuhannya secara online, hal ini sejalan dengan penerapan kebijakan pemerintah untuk bekerja dari rumah atau *Work From Home (WFH)* dan sekolah dari rumah. Semakin majunya *ecommerce* di Indonesia berdampak meledaknya jumlah pengguna jasa pengiriman serta logistik, Banyak penjual yang melaksanakan transaksi di *platform e-commerce*, Demikian pula dengan organisasi yang bergerak dalam bidang jasa ekspedisi. Telah pasti teknologi informasi berfungsi bernilai, dimana tiap barang yang hendak dikirim wajib dikontrol berbasis *teknologi informasi*, mulai dari barang diserahkan oleh pengirim ke petugas ekspedisi, setelah itu di catat, hingga barang sampai ke alamat tujuan[2].

Ecommerce di Indonesia sangat berperan penting dalam pengguna jasa pengiriman. Jasa ekspedisi mampu mengantarkan ke daerah, domestik maupun luar negeri, untuk dapat berkontribusi dalam proses pengiriman. Jasa kurir juga berperan penting dalam menunjang kelancaran bisnis perusahaan yang membutuhkan jasa pengiriman yang cepat dan aman. Ada banyak perusahaan ekspedisi yang populer di Indonesia misalnya jasa ekspedisi pertama dan JASA EKSPEDISI KEDUA. Perusahaan ekspedisi gencar sediakan media komunikasi dengan pelanggan. Salah satunya ialah dengan membagikan pelayanan berbentuk konsultasi, tanya jawab seputar jasa pengiriman tersebut dengan memanfaatkan teknologi informasi berbentuk media sosial. Twitter, Twitter digunakan 4 selaku fasilitas buat promosi produk, iklan, kampanye politik ataupun selaku fasilitas menyatakan komentar terkait kritik, masukan, isu- isu serta opini- opini publik. Twitter dikira lebih diminati oleh para warga Indonesia karna dirasa lebih gampang serta sederhana dalam merepresentasikan opininya. Analisis terhadap Twitter yang dikirim, ber-guna bagi perusahaan untuk meningkatkan *performa* layanan. penilaian buruk dari pelanggan karena keterlambatan pengiriman ekspedisi. Masalah tersebut dapat diatasi dengan menganalisa ulasan atau sentimen pengguna Twitter terhadap akun jasa ekspedisi untuk menghasilkan informasi mengenai *performa* perusahaan. Pada penelitian ini data yang digali berkaitan tentang sentimen pengguna jasa ekspedisi di Indonesia pada akun media sosial Twitter @jasa ekspedisi pertama ID, dan @jasa ekspedisi kedua.id. Pemilihan kedua jasa ekspedisi tersebut didasari aktifnya akun Twitter sehingga memudahkan pengambilan dari data. Peneliti juga membandingkan ulasan positif maupun negatif

dari pemain baru Jasa ekspedisi kedua dan pemain lama yaitu jasa ekspedisi pertama. Dalam pengklasifikasian akan digunakan model *naive bayes* dan KNN, kedua model ini digunakan karena mudah dalam implementasinya. Pengujian perbandingan akurasi dari kedua model tersebut adalah untuk mengetahui akurasi terbaik dalam menyelesaikan *klasifikasi* dengan data teks sentimen jasa ekspedisi. *Naive bayes* adalah model yang sangat efektif digunakan dalam penyelesaian masalah *klasifikasi*[3].

Penelitian ini menggunakan *Naive Bayes* serta KNN (K-Nearest Neighbor) menggunakan dataset sentimen jasa ekspedisi pertama dan Jasa ekspedisi kedua. Hasil disimpulkan bahwa eksperimen dengan menggunakan *Labeling* otomatis lebih *representatif* dalam menentukan hasil terbaik kelas data. Model *percentage split 70:30* memiliki hasil akurasi terbaik pada Data yang diuji eksperimen jasa ekspedisi pertama dan Jasa ekspedisi kedua. studi kasus jasa ekspedisi pertama memiliki nilai akurasi 97% untuk algoritma KNN, studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 89% untuk algoritma *Naive Bayes* pada *Labeling otomatis*.

Permasalahan pada perusahaan ekspedisi jasa ekspedisi pertama dan jasa ekspedisi kedua mendapatkan penilaian buruk dari pelanggan karena keterlambatan pengiriman ekspedisi masalah tersebut diatasi dengan menganalisa ulasan atau sentimen pengguna di twitter. Penelitian ini menggunakan metode algoritma *naive bayes* dan algoritma *k-nearest neighbor* untuk menganalisis dan *klasifikasi* ulasan terhadap jasa pengiriman ekspedisi.

II. LITERATUR REVIEW

A. Penelitian Terkait

Salam, J. Zeniarja, and R. S. U. Khasanah Pengujian[4] *klasifikasi* Analisa sentimen dokumen komentar pada jasa ekspedisi barang (J&T) di facebook menggunakan algoritma KNN (K- Nearest Neighbor) yang memiliki *accuracy* tertinggi yaitu 79,21% dengan jumlah dokumen training 101 sedangkan *accuracy* terendahnya adalah 70,3% dengan jumlah dokumen training 101 A.

F. N. Zuhri and A. Alamsyah[5].*Klasifikasi Analisis* sentimen terhadap data Kaskus mengenai brand Smartfren dilakukan dengan metode *Naive Bayes*, dengan menggunakan 1000 data latih, didapatkan *precision* sebesar 98.42 %, *recall* sebesar 98.40 %, dan tingkat akurasi sebesar 98.40 %. Hal ini menunjukkan bahwa sistem dapat memisahkan data bersentimen positif dan negatif dengan sangat baik. Dan juga memperoleh nilai koefisien *kappa* sebesar 0.948, maka menunjukkan kesepakatan yang sangat kuat antar penguji karena mendekati nilai 1, maka instrumen data latih dapat dikatakan reliabel. Respon atau sentimen masyarakat terhadap brand Smartfren di Kaskus menggunakan data uji sebanyak 6388 data. Hasil *klasifikasi* menunjukkan 4049 data masuk ke dalam sentimen positif dan hanya 2339 data masuk ke dalam sentimen negatif.

Ramadhani, Slamet Harry, Wahyudin, Muhammad Iwan[6]. Penelitian didapati tingkat keakurasian yang

berbeda, untuk penggunaan metode Naïve Bayes menghasilkan tingkat akurasi sebesar 90.71% +/- 4.66% (*micro average*: 90.77%) sedangkan untuk metode KNN menghasilkan tingkat akurasi sebesar 74.78% +/- 3.74% (*micro average*: 74.77%)

M. Syarifuddin[7]Klasifikasi metode Naïve Bayes dan KNN, serta mengetahui kecenderungan opini masyarakat di twitter. Subjek diambil menggunakan API Twitter sebanyak 1098 opini dengan kata kunci “COVID-19”, Dimulai dari klasifikasi opini positif atau negatif, data *cleansing*, *preprocessing*, hingga didapatkan hasil akhir, memiliki nilai lebih tinggi, salah satunya adalah tingkat *accuracy* sebesar 63.21%, sedangkan metode KNN sebesar 58.10%, dan didapatkan pula kecenderungan opini masyarakat di twitter condong positif, hal tersebut dapat dilihat dari jumlah opini positif sebesar 610

M. Nur Habibi and Sunjana[8]Klasifikasi emosi positif 91,27%, 7,56% Sentimen negatif dan sentimen netral 1,17%. ini Klasifikasi tersebut menghasilkan akurasi rata-rata 69,2%. Emosi Jokowi dan emosi Prabowo 100%. Hitung akurasi klasifikasi menggunakan ROC metode. Hasil akhir dari analisis jaringan sosial berdasarkan - Perhitungan properti jaringan menghasilkan 277 node, 7.950 tepi, rata-rata 57.401 derajat, rata-rata 56,44 Derajat tertimbang, dengan diameter jaringan 4 dan rata-rata 1,853 Panjang jalur, kepadatan 0,201, dan jumlah komunitas 5.

R. N. Devita, H. W. Herwanto, and A. P. Wibawa[9] Menunjukkan metode Naïve Bayes memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan metode K-Nearest Neighbor memiliki tingkat akurasi yang cukup rendah yaitu 40%.

B. Algoritma Klasifikasi

Metode Klasifikasi adalah paling sederhana dan paling umum digunakan. Model klasifikasi Naive Bayes menghitung *probabilitas posterior* suatu kelas berdasarkan distribusi kata dalam dokumen. Hal itu bergantung pada *representasi* dokumen yang sangat sederhana sebagai *Bag of words*. Model ini bekerja dengan mengekstraksi *fitur bag of words* yang mengabaikan posisi kata dalam dokumen[10].

Naïve Bayes Classifier (NBC) yaitu proses analisis terhadap sampel dokumen berupa pemilihan *term*, yaitu kata yang mungkin muncul dalam setiap dokumen sampel yang sedapat mungkin dapat menjadi *representasi* dokumen. Kemudian penentuan *probabilitas prior* bagi setiap kategori berdasarkan sampel dokumen, metode *probabilitas* dan *statistic* yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Menurut *teorema* Bayes. Pada tahap pengujian, ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi pada *probabilitas*. Naïve Bayes Classifier dalam melakukan klasifikasi terdapat dua proses penting yaitu *learning (training)* dan *testing*[5].

K-Nearest Neighbor (K-NN) adalah algoritma ini digunakan untuk penelitian suatu klasifikasi dokumen komentar pada situs youtube dengan menggunakan algoritma K-Nearest Neighbor (K-NN) oleh Moh Aziz Nugroho, membahas tentang klasifikasi dokumen komentar pada situs youtube menggunakan algoritma K-Nearest Neighbor (K-NN). Dalam penelitiannya diperoleh data *review* film berbahasa Inggris ke dalam dua Sentiment yaitu positif dan negatif pada *review* film atau *trailer* bergenre aksi disitus youtube.com Sebuah metode *lazy learning* yang dimana tidak ada model yang dipelajari dari data pelatihan, yang terjadi hanya belajar bila contoh uji harus diklasifikasikan. KNN (K-Nearest Neighbor) bertujuan untuk mengklasifikasi objek berdasarkan *atribut* dan data *training*. Data klasifikasi tidak menggunakan data apapun untuk dicocokkan dan hanya berdasarkan memori. Didalam *query* terdapat titik yang akan ditemukan sejumlah k titik *training* yang posisinya paling dekat dengan titik *query*[4]

III. DATASET

Pada penelitian ini data disimpan dengan *format .csv* dengan total *record* data untuk studi kasus jasa ekspedisi pertama sebanyak 2284 pada Table I dan Jasa ekspedisi kedua Sebanyak . Sedangkan total *record* data untuk studi kasus Jasa ekspedisi kedua Sebanyak 2340 pada Table II. Berikut merupakan contoh dari dataset dari dua studi kasus:

Table I. Dataset jasa ekspedisi pertama

NO	jasa ekspedisi pertama	
	Query Twitter	Jumlah Data
1	jasa ekspedisi pertama	351
2	jasa ekspedisi pertama Id	472
3	jasa ekspedisi pertama Care	461
4	Kurir jasa ekspedisi pertama	300
5	Agen jasa ekspedisi pertama	350
6	jasa ekspedisi pertama Ekpress	350
Total Data		2284

Table II. Dataset Jasa ekspedisi kedua

NO	Jasa ekspedisi kedua	
	Query Twitter	Jumlah Data
1	Jasa ekspedisi kedua	740
2	Jasa ekspedisi kedua Id	550
3	Kurir Jasa ekspedisi kedua	350
4	Dijasa ekspedisi kedua	175

NO	Jasa ekspedisi kedua	
	Query Twitter	Jumlah Data
5	Jasa ekspedisi kedua lama	275
6	Ekspedisi Jasa ekspedisi kedua	250
Total Data		2340

A. Preprocessing Data

Preprocessing data adalah teks seperti case folding, membersihkan dataset, tokenisasi, normalisasi, stopword removal, dan stemming. Pembersihan data dilakukan untuk eliminasi tanda baca, simbol, angka, emoticon pada keseluruhan dokumen. dari data mentah menjadi data yang siap pakai agar memudahkan proses klasifikasi menjadi positif dan negatif. Data pre-processing adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Pada penelitian ini digunakan lima langkah dalam tahapan preprocessing. Berikut adalah penjelasan dari lima langkah preprocessing:

- Cleansing Data** : Cleansing dalam penelitian ini bertujuan untuk menghapus RT, username, hashtag, angka dan URL pada tweets.
 Contoh tweet : Malem teh pengen santai maen game abis kuliah seharian. Ehhhh udah 2 jam internet mati. Gimana nih @IndiHome
 Hasil cleansing : Malem teh pengen santai maen game abis kuliah seharian. Ehhhh udah jam internet mati. Gimans nih?
- Case Folding** : Case Folding dalam penelitian ini bertujuan untuk merubah teks dari penulisan huruf besar menjadi penulisan huruf kecil.
- Hasil case folding** : malem teh pengen santai maen game abis kuliah seharian. ehhhh udah jam internet mati. gimana nih
- Tokenization** : Tokenization dalam penelitian ini bertujuan untuk memisahkan kata kata dari dalam kalimat.
 Hasil tokenization : malem, teh, pengen, santai, maen, game, abis, kuliah, seharian, ehhhh, udah, jam, internet, mati, gimana, nih,
- Stopword Removal** : Stopword Removal dalam penelitian ini bertujuan untuk menghapus kata sambung
 Contoh dari Stopword Removal bahasa Indonesia adalah sebagai berikut : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, pada, dengan, adalah, yaitu, ke, tak, tidak, di, pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, dll [11]
 Hasil stopword removal: malem, teh, santai, maen, game, abis, kuliah, seharian, ehhhh, udah, jam, internet, mati, gimana, nih,
- Stemming**: Stemming dalam penelitian ini bertujuan untuk membentuk kata kata dasar.
 Hasil stemming: malem, teh, santai, maen, game, abis, kuliah, hari, ehhhh, udah, jam, internet, mati, gimana, nih,

- Remove Duplicate**: Remove Duplicate dalam penelitian ini bertujuan untuk menghapus data yang memiliki duplikat.

Setelah melakukan tahap *preprocessing*, data yang telah 1817 pada Gambar 1, *dipreprocessing* menjadi 1 untuk data jasa ekspedisi pertama dan 1379 untuk data Jasa ekspedisi kedua. Berikut merupakan hasil data setelah dilakukan tahapan *preprocessing*:

	Klasifikasi	Tweet
2028	Negatif	
2202	Negatif	a thread barang bom sampe aduhhh cape banget p...
1929	Positif	aaku kaa
2170	Negatif	abis kepo twitter yg complain jne degan ngrim...
1957	Positif	adaa netflix k k k k k user k user k payment a...
...
456	Positif	yuk submit video banyak dapat hadiah oiya dapa...
1928	Positif	yukk kak
1982	Positif	yukk kakk nonton rowoon pacar netflix sharing ...
1994	Positif	yuu ridi netflix no vpn sharing k k k pu k pu ...
1963	Positif	yuuuk haloo cari app premium murahh yuk full g...

1817 rows x 2 columns

Gambar 1. *Preprocessing* jasa ekspedisi pertama dengan Labeling

	Klasifikasi	Tweet
1033	Positif	
1491	Positif	abang anteraja gercep bgt gak ngepacking udah ...
1339	Positif	abee yg udh tua
290	Positif	ada paket anteraja serius tanggap complain jua...
2087	Positif	ada paket kirim nyasar hub surabaya note ya ga ...
...
2246	Positif	yg pakai anteraja gak kayak gimana ekspedisi e...
477	Positif	yg waras j l id express anteraja sih
841	Negatif	you re so not professional lambat kemarin mala ...
434	Positif	yuk yuk cepet kirim
1284	Positif	yukkk bisaaa yukkk

1379 rows x 2 columns

Gambar 2. *Preprocessing* Jasa ekspedisi kedua dengan Labeling

	Tweet
2028	
2202	a thread barang bom sampe aduhhh cape banget p...
1929	aaku kaa
2170	abis kepo twitter yg komplain jne degan ngirim...
1957	adaa netflix k k k k k user k user k payment a...
...	...
456	yuk submit video banyak dapat hadiah oiya dapa...
1928	yukk kak
1982	yukk kakk nonton rowoon pacar netflix sharing ...
1994	yuu ridi netflix no vpn sharing k k k pu k pu ...
1953	yuuuk haloo cari app premium murahh yuk full g...

1818 rows × 1 columns

Gambar 3. *Preprocessing* jasa ekspedisi pertama tanpa *Labeling* *Tweet*

1033	
1491	abang anteraja gercep bgt gak ngepacking udah ...
1339	abee yg udh tua
290	ada paket anteraja serius tanggap komplain jua...
2087	ada paket kirim nyasar hub surabaya note ya ga...
...	...
2245	yg pakai anteraja gak kayak gimana ekspedisi e...
477	yg waras j t id express anteraja sih
841	you re so not professional lambat kemarin mala...
434	yuk yuk cepet kirim
1284	yukkk bisaaa yukkk

1379 rows × 1 columns

Gambar 4. *Preprocessing* Jasa ekspedisi kedua tanpa *Labeling*

B. Labeling

pada alur ini, dataset yang sudah dibersihkan pada saat preprocessing akan dilakukan pelabelan kelas data. Pelabelan kelas data dibagi menjadi dua, kelas positif dan kelas negatif. Proses pelabelan dilakukan dengan menggunakan bahasa R melalui Tools Google Colab. Proses pelabelan dengan menggunakan bahasa R diperlukan kamus bahasa Indonesia yang terintegrasi dengan Google Drive. Kamus berbahasa Indonesia yang digunakan terdiri dari dua kamus, yaitu kamus berbahasa Indonesia dengan kata positif dan kamus berbahasa Indonesia dengan kata Negatif[5].

Selanjutnya labeling, dilakukan *skoring* pada tiap kalimat. Setiap kata yang terdeteksi diberikan skoring untuk menilai kelas sentimen. Untuk kata positif, diberikan nilai 1 sedangkan untuk kata negatif diberikan nilai -1. Jika terdapat kata yang tidak ada dalam kamus positif ataupun negatif, diberikan nilai 0. Pemberian *skoring* dilakukan dengan cara menghitung jumlah nilai pada tiap kata dalam satu kalimat. Jika nilai ≥ 0 maka dilabeli sebagai sentimen *tweet* positif, sebaliknya jika nilai < 0 maka dilabeli sebagai *tweet* sentimen negatif. Jika hasil skoring bernilai 0 maka dilabeli sebagai

tweet positif. Kelas data positif akan diberi nilai 1, sedangkan kelas data Negatif akan diberi nilai 0.

Pada proses Lebeling positif otomatis dan labeling negatif otomatis memiliki kosa kata yang sudah di rangkum dalam .txt bernama corpus positif dan corpus negatif untuk mendeteksi kata demi kata disetiap narasi opini masyarakat tweets. Table III.

Table III. Nilai Klasifikasi Otomatis

Klasifikasi Otomatis		
Sentimen	Positif	Negatif
jasa ekspedisi pertama	732	646
Jasa ekspedisi kedua	1005	812

Pada penelitian ini juga dilakukan eksperimen dengan menggunakan labeling manual terhadap kedua dataset tersebut. Label positif merupakan *tweet* yang memiliki pujian sedangkan *tweet* negative berupa kata-kata kasar dan sarkas Table IV.

Table IV. Klasifikasi Manual

Klasifikasi Manual		
Sentimen	Positif	Negatif
jasa ekspedisi pertama	1135	683
Jasa ekspedisi kedua	912	467

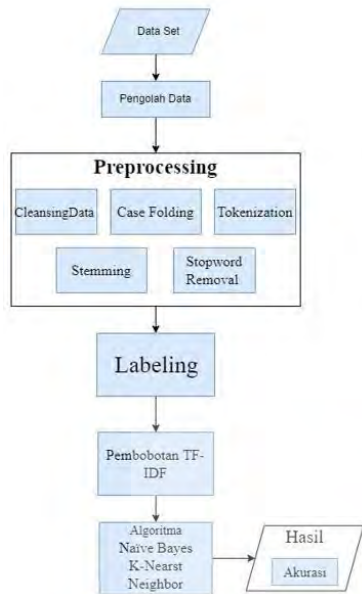
C. TF-IDF

Data hasil *preprocessing* yang berupa kata akan diubah ke dalam bentuk angka dengandilakukan proses pembobotan kata yang bertujuan untuk menghitung bobot pada masing-masing kata yang akan digunakan sebagai fitur, semakin banyak dokumen yang akan diproses maka semakin banyak fitur. Pada tahapanini terdapat dua bagian proses yaitu *TF* (*Term Frequency*) dan *IDF* (*InverseDocument Frequency*), *TF* adalah jumlah kemunculan tiap bobotlebih kecil apabila kata tersebut sering muncul pada setiapdokumen dalam koleksi, sebaliknyabobot *TF-IDF* akan lebih besar apabilakata tersebut jarang muncul pada setiap dokumendalam koleksi. Dalam penelitian ini pembobotan *TF-IDF* yang digunakan adalah *TF-IDF* tanpa normalisasi[12]

IV. METODE

The penelitian ini tercantum dalam sifat *kuantitatif* ataupun mempunyai *prosedur* penelitian kuantitatif. Jenis penelitian ini memanfaatkan angka- angka ataupun mengumpulkan informasi yang mendalam serta lengkap buat menggambarkan *fenomena* ataupun kenyataan penelitian. Pengumpulan data dilakukan menggunakan objek penelitian secara langsung. Dimana penelitian ini mengambil data informasi tweet dari twitter yang bisa intergrasikan Twitter API menggunakan *tools* Rapid Miner penelitian ini *python* dan *r tools* dibantu untuk melakukan Preprocessing dan labeling pada *Google Colab*. Dalam Penelitian ini sedang di

uji eksperimen perbandingan kinerja algoritma klasifikasi Naïve Bayes dan K-Nearest Neighbor untuk mengetahui algoritma memiliki tingkat akurasi tinggi. Tahapan eksperimen menjelaskan tentang proses mengumpulkan data sampai pada tahap evaluasi dan *validasi* hasil, sehingga dapat menghasilkan hasil perbandingan dari masing-masing algoritma berikut ilustrasikan pada Gambar 3.



Gambar 3. Tahapan Penelitian

Tahapan awal melakukan yaitu mengumpulkan dataset-dataset. tahapan berikutnya dilakukan tahapan *preprocessing* untuk membersihkan dataset mentah. Setelah data selesai dilakukan *preprocessing*, tahapan berikutnya adalah dilakukan labeling untuk mengetahui kelas positif dan negatif. Setelah itu dilakukan tahapan pembobotan *TF-IDF* untuk memberikan bobot pada setiap kata yang terdapat di dalam dokumen. Selanjutnya, dilakukan metode klasifikasi dengan menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor.

V. HASIL SEMUA EKSPERIMEN

Akan menampilkan hasil semua eksperimen yang telah dilakukan. Pada penelitian ini melakukan eksperimen pada dua studi kasus ekspedisi yaitu jasa ekspedisi pertama dan Jasa ekspedisi kedua. Eksperimen dilakukan dengan menggunakan dua algoritma yaitu Naïve Bayes dan KNN (K-Nearest Neighbor). Terdapat tiga *skenario* yang menjadi pembagian *persentase* data. Pembagian presentasi data merupakan pemisahan data *training* dan *testing* berdasarkan persentasenya, contoh 90% : 10% berarti 90% merupakan data *training* dan 10% merupakan data *testing*. Pada penelitian ini pembagian *persentase* dibagi menjadi tiga skenario eksperimen yaitu eksperimen pertama menggunakan pembagian data 90% : 10%, eksperimen kedua menggunakan pembagian data 80% : 20% dan eksperimen yang ketiga menggunakan pembagian data 70% : 30%. Penelitian ini juga membandingkan nilai akurasi yang

dilakukan dengan menggunakan metode labeling manual dan metode labeling otomatis

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau *percentage split* 90% : 10% bahwa untuk studi kasus jasa ekspedisi pertama memiliki nilai akurasi 90% untuk algoritma KNN pada gambar 4 dan 78% untuk algoritma Naïve Bayes Pada gambar 5. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 85% untuk algoritma KNN pada gambar 6 dan 89% untuk algoritma Naïve Bayes pada gambar 7. Berikut merupakan hasil dari kinerja *percentage split* 90% : 10%.

```

0.8994708994708994
[[ 54 16]
 [ 3 116]]
precision  recall  f1-score  support
0          0.95    0.77     0.85     70
1          0.88    0.97     0.92    119
accuracy
macro avg  0.91    0.87     0.89    189
weighted avg  0.90    0.90     0.90    189
  
```

Gambar 4. KNN jasa ekspedisi pertama *Split* 90:10 *Labeling Otomatis*

```

0.783068783068783
[[56 14]
 [27 92]]
precision  recall  f1-score  support
0          0.67    0.80     0.73     70
1          0.87    0.77     0.82    119
accuracy
macro avg  0.77    0.79     0.77    189
weighted avg  0.80    0.78     0.79    189
  
```

Gambar 5. Naïve Bayes jasa ekspedisi pertama *Split* 90:10 *Labeling Otomatis*

```

0.8497109826589595
[[ 47 24]
 [ 2 100]]
precision  recall  f1-score  support
0          0.96    0.66     0.78     71
1          0.81    0.98     0.88    102
accuracy
macro avg  0.88    0.82     0.83    173
weighted avg  0.87    0.85     0.84    173
  
```

Gambar 6. KNN Jasa ekspedisi kedua *Split* 90:10 *Labeling Otomatis*

```

0.8901734104046243
[[56 15]
 [ 4 98]]
precision  recall  f1-score  support
0          0.93    0.79     0.85     71
1          0.87    0.96     0.91    102
accuracy
macro avg  0.90    0.87     0.88    173
weighted avg  0.89    0.89     0.89    173
  
```

Gambar 7. Naïve Bayes Jasa ekspedisi kedua *Split* 90:10 *Labeling Otomatis*

Hasil dari kinerja model dengan menggunakan metode *labeling otomatis* dan pembagian data atau *percentage split* 80% : 20% bahwa untuk studi kasus jasa ekspedisi pertama

memiliki nilai akurasi 86% untuk algoritma KNN pada gambar 8 dan 80% untuk algoritma Naïve Bayes pada gambar 9. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 97% untuk algoritma KNN pada gambar 10 dan 81% untuk algoritma Naïve Bayes pada gambar 11. Berikut merupakan hasil dari kinerja *percentage split 80% : 20%*.

```

0.8571428571428571
[[ 56 14]
 [ 13 106]]
precision  recall  f1-score  support
0          0.81    0.80    0.81     70
1          0.88    0.89    0.89    119

accuracy
macro avg  0.85    0.85    0.85    189
weighted avg 0.86    0.86    0.86    189

```

Gambar 8. KNN jasa ekspedisi pertama *Split 80:20 Labeling Otomatis*

```

0.8042328042328042
[[118 36]
 [ 38 186]]
precision  recall  f1-score  support
0          0.76    0.77    0.76    154
1          0.84    0.83    0.83    224

accuracy
macro avg  0.80    0.80    0.80    378
weighted avg 0.80    0.80    0.80    378

```

Gambar 9. Naïve Bayes jasa ekspedisi pertama *Split 80:20 Labeling Otomatis*

```

0.9710982658959537
[[ 68  3]
 [  2 100]]
precision  recall  f1-score  support
0          0.97    0.96    0.96     71
1          0.97    0.98    0.98    102

accuracy
macro avg  0.97    0.97    0.97    173
weighted avg 0.97    0.97    0.97    173

```

Gambar 10. KNN Jasa ekspedisi kedua *Split 80:20 Labeling Otomatis*

```

0.8057971014492754
[[ 79 45]
 [ 22 199]]
precision  recall  f1-score  support
0          0.78    0.64    0.70    124
1          0.82    0.90    0.86    221

accuracy
macro avg  0.80    0.77    0.78    345
weighted avg 0.80    0.81    0.80    345

```

Gambar 11. Naïve Bayes Jasa ekspedisi kedua *Split 80:20 Labeling Otomatis*

Hasil dari kinerja model dengan menggunakan metode *Labeling* otomatis dan pembagian data atau *percentage split 70% : 30%* bahwa untuk studi kasus jasa ekspedisi pertama memiliki nilai akurasi 86% untuk algoritma KNN pada gambar12 dan 85% untuk algoritma Naïve Bayes pada gambar 13. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 97% untuk algoritma KNN pada gambar14 dan 78% untuk algoritma Naïve Bayes pada gambar 15. Berikut merupakan hasil dari kinerja *percentage split 70%: 30%*.

```

0.8571428571428571
[[ 56 14]
 [ 13 106]]
precision  recall  f1-score  support
0          0.81    0.80    0.81     70
1          0.88    0.89    0.89    119

accuracy
macro avg  0.85    0.85    0.85    189
weighted avg 0.86    0.86    0.86    189

```

Gambar 12. KNN jasa ekspedisi pertama *Split 70:30 Labeling Otomatis*

```

0.8465608465608465
[[207 41]
 [ 46 273]]
precision  recall  f1-score  support
0          0.82    0.83    0.83    248
1          0.87    0.86    0.86    319

accuracy
macro avg  0.84    0.85    0.84    567
weighted avg 0.85    0.85    0.85    567

```

Gambar 13. Naïve Bayes jasa ekspedisi pertama *Split 70:30 Labeling Otomatis*

```

0.9710982658959537
[[ 66  5]
 [  0 102]]
precision  recall  f1-score  support
0          1.00    0.93    0.96     71
1          0.95    1.00    0.98    102

accuracy
macro avg  0.98    0.96    0.97    173
weighted avg 0.97    0.97    0.97    173

```

Gambar 14. KNN Jasa ekspedisi kedua *Split 70:30 Labeling Otomatis*

```

0.7794970986460348
[[ 99 91]
 [ 23 304]]
precision  recall  f1-score  support
0          0.81    0.52    0.63    190
1          0.77    0.93    0.84    327

accuracy
macro avg  0.79    0.73    0.74    517
weighted avg 0.79    0.78    0.77    517

```

Gambar 15. Naïve Bayes Jasa ekspedisi kedua *Split 70:30 Labeling Otomatis*

Hasil dari kinerja model dengan menggunakan metode *labeling* manual dan pembagian data atau *percentage split 90% : 10%* bahwa untuk studi kasus jasa ekspedisi pertama memiliki nilai akurasi 64% untuk algoritma KNN pada gambar 16 dan 68% untuk algoritma Naïve Bayes pada gambar 17. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 66% untuk algoritma KNN pada gambar 18 dan 60% untuk algoritma Naïve Bayes pada gambar19. Berikut merupakan hasil dari kinerja *percentage split 90% : 10%*.

```

0.6363636363636364
[[29 43]
 [ 5 55]]
precision  recall  f1-score  support
0          0.85    0.40    0.55     72
1          0.56    0.92    0.70     60

accuracy
macro avg  0.71    0.66    0.62    132
weighted avg 0.72    0.64    0.61    132

```

Gambar 16. KNN jasa ekspedisi pertama *Split 90:10 Labeling Manual*

```

-----
0.6818181818181818
[[59 13]
 [29 31]]
precision recall f1-score support
0 0.67 0.82 0.74 72
1 0.70 0.52 0.60 60

accuracy 0.68 132
macro avg 0.69 0.67 0.67 132
weighted avg 0.69 0.68 0.67 132

```

Gambar 17. Naïve Bayes jasa ekspedisi pertama *Split 90:10 Labeling Manual*

```

-----
0.6636363636363637
[[16 35]
 [ 2 57]]
precision recall f1-score support
0 0.89 0.31 0.46 51
1 0.62 0.97 0.75 59

accuracy 0.66 110
macro avg 0.75 0.64 0.61 110
weighted avg 0.74 0.66 0.62 110

```

Gambar 18. KNN Jasa ekspedisi kedua *Split 90:10 Labeling Manual*

```

-----
0.6
[[ 9 42]
 [ 2 57]]
precision recall f1-score support
0 0.82 0.18 0.29 51
1 0.58 0.97 0.72 59

accuracy 0.60 110
macro avg 0.70 0.57 0.51 110
weighted avg 0.69 0.60 0.52 110

```

Gambar 19. Naïve Bayes Jasa ekspedisi kedua *Split 90:10 Labeling Manual*

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau *percentage split* 80% : 20% bahwa untuk studi kasus jasa ekspedisi pertama memiliki nilai akurasi 90% untuk algoritma KNN pada gambar 20 dan 71% untuk algoritma Naïve Bayes pada gambar 21. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 94% untuk algoritma KNN pada gambar 22 dan 60% untuk algoritma Naïve Bayes pada gambar 23. Berikut merupakan hasil dari kinerja *percentage split* 80% : 20%.

```

-----
0.9015151515151515
[[66 6]
 [ 7 53]]
precision recall f1-score support
0 0.90 0.92 0.91 72
1 0.90 0.88 0.89 60

accuracy 0.90 132
macro avg 0.90 0.90 0.90 132
weighted avg 0.90 0.90 0.90 132

```

Gambar 20. KNN jasa ekspedisi pertama *Split 80:20 Labeling Manual*

```

-----
0.7083333333333334
[[122 20]
 [ 57 65]]
precision recall f1-score support
0 0.68 0.86 0.76 142
1 0.76 0.53 0.63 122

accuracy 0.71 264
macro avg 0.72 0.70 0.69 264
weighted avg 0.72 0.71 0.70 264

```

Gambar 21. Naïve Bayes *Split* jasa ekspedisi pertama *80:20 Labeling Manual*

```

-----
0.9363636363636364
[[48 3]
 [ 4 55]]
precision recall f1-score support
0 0.92 0.94 0.93 51
1 0.95 0.93 0.94 59

accuracy 0.94 110
macro avg 0.94 0.94 0.94 110
weighted avg 0.94 0.94 0.94 110

```

Gambar 22. KNN Jasa ekspedisi kedua *Split 80:20 Labeling Manual*

```

-----
0.5954545454545455
[[ 23 77]
 [ 12 108]]
precision recall f1-score support
0 0.66 0.23 0.34 100
1 0.58 0.90 0.71 120

accuracy 0.60 220
macro avg 0.62 0.57 0.52 220
weighted avg 0.62 0.60 0.54 220

```

Gambar 23. Naïve Bayes Jasa ekspedisi kedua *Split 80:20 Labeling Manual*

Hasil dari kinerja model dengan menggunakan metode labeling manual dan pembagian data atau *percentage split* 70% : 30% bahwa untuk studi kasus jasa ekspedisi pertama memiliki nilai akurasi 94% untuk algoritma KNN pada gambar 24 dan 67% untuk algoritma Naïve Bayes pada gambar 25. Sedangkan untuk studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 95% untuk algoritma KNN pada gambar 26 dan 57% untuk algoritma Naïve Bayes pada gambar 27. Berikut merupakan hasil dari kinerja *percentage split* 70% : 30%.

```

-----
0.9393939393939394
[[66 6]
 [ 2 58]]
precision recall f1-score support
0 0.97 0.92 0.94 72
1 0.91 0.97 0.94 66

accuracy 0.94 132
macro avg 0.94 0.94 0.94 132
weighted avg 0.94 0.94 0.94 132

```

Gambar 24. KNN jasa ekspedisi pertama *Split 70:30 Labeling Manual*

```

-----
0.6658227848101266
[[163 35]
 [ 97 100]]
precision recall f1-score support
0 0.63 0.82 0.71 198
1 0.74 0.51 0.60 197

accuracy 0.67 395
macro avg 0.68 0.67 0.66 395
weighted avg 0.68 0.67 0.66 395

```

Gambar 25. Naïve Bayes jasa ekspedisi pertama *Split 70:30 Labeling Manual*

```

-----
0.9545454545454546
[[46 5]
 [ 0 59]]
precision recall f1-score support
0 1.00 0.90 0.95 51
1 0.92 1.00 0.96 59

accuracy 0.95 110
macro avg 0.96 0.95 0.95 110
weighted avg 0.96 0.95 0.95 110

```

Gambar 26. KNN Jasa ekspedisi kedua *Split 70:30 Labeling Manual*

```

0.5666666666666667
[[ 32 112]
 [ 31 155]]
precision    recall  f1-score   support

   0         0.51    0.22    0.31     144
   1         0.58    0.83    0.68     186

accuracy          0.57    330
macro avg         0.54    0.53    0.50    330
weighted avg      0.55    0.57    0.52    330

```

Gambar 27. Naïve Bayes Jasa ekspedisi kedua *Split 70:30 Labeling Manual*

Hasil dari penelitian ini adalah eksperimen dengan menggunakan algoritma K-NEAREST NEIGHBOR memiliki nilai akurasi paling tinggi pada tiga skenario eksperimen percentage *split* pada dua model data yang diuji yaitu jasa ekspedisi pertamadan Jasa ekspedisi kedua dengan menggunakan *labeling otomatis* dan manual. Hasil dari pembahasan terangkum dalam tabel beri

Table V. Perbandingan Akurasi *Labeling Otomatis*

Percobaan	jasa ekspedisi pertama		Jasa ekspedisi kedua	
	KNN	NB	KNN	NB
90:10	90%	78%	85%	89%
80:20	86%	80%	97%	81%
70:30	86%	85%	97%	78%

Table VI. Perbandingan Akurasi *Labeling Manual*

Percobaan	jasa ekspedisi pertama		Jasa ekspedisi kedua	
	KNN	NB	KNN	NB
90:10	64%	68%	66%	60%
80:20	90%	71%	94%	60%
70:30	94%	67%	95%	57%

VI. KESIMPULAN

Penelitian ini menggunakan algoritma Naïve Bayes serta KNN (K- Nearest Neighbor) menggunakan dataset sentimen jasa ekspedisi jasa ekspedisi pertama dan Jasa ekspedisi kedua. berikut disimpulkan uji eksperimen dengan menggunakan labeling otomatis lebih representatif dalam menentukan kelas data. Model percentage split 70:30 memiliki hasil akurasi terbaik pada Data yang diuji eksperimen jasa ekspedisi pertama dan Jasa ekspedisi kedua. studi kasus jasa ekspedisi pertama memiliki nilai akurasi 86% untuk algoritma KNN (K- Nearest Neighbor), studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 97% untuk algoritma KNN (K- Nearest Neighbor) pada eksperimen Labeling otomatis. Pada Labeling manual model percentage split 70:30 studi kasus jasa ekspedisi pertama memiliki nilai akurasi 94% untuk algoritma KNN (K- Nearest Neighbor), studi kasus Jasa ekspedisi kedua memiliki nilai akurasi 95% K- Nearest Neighbor. Bahwa hasil perhitungan menunjukkan metode KNN(K- Nearest Neighbor)) Labeling otomatis

studi kasus Jasa ekspedisi kedua memiliki hasil tingkat akurasi yang lebih tinggi dengan nilai rata-rata akurasi sebesar 97% yang memiliki nilai rata-rata paling baik dari penelitian ini.

REFERENCES

- [1] M. (Muhammad) Rifqi and N. K. (Nia) Wardhani, "Aplikasi Peran dan Kegunaan Teknologi Near Field Communication (Nfc) terhadap Kegiatan Proses Belajar Mengajar di Perguruan Tinggi," *J. Ilmu Tek. dan Komput.*, vol. 1, no. 1, pp. 20–26, 2017, [Online]. Available: <https://www.neliti.com/publications/237558/>.
- [2] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.
- [3] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Pros. SNST*, vol. 7, pp. 37–42, 2016, [Online]. Available: https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/1506/1589.
- [4] A. Salam, J. Zeniarja, and R. S. U. Khasanah, "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)," *Pros. SINTAK*, pp. 480–486, 2018.
- [5] F. N. Zuhri and A. Alamsyah, "Menggunakan Naïve Bayes Classifier Di Forum Kaskus Publik Sentiment Analysis Of Smartfren Brand Using Naïve Bayes Classifier On Kaskus Forum," *e- Proceeding Manag.*, vol. 4, no. 1, pp. 242–251, 2017.
- [6] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022, doi: 10.35870/jtik.v6i4.530.
- [7] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan KNN," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [8] M. Nur Habibi and Sunjana, "Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 11, pp. 22–30, 2019, doi: 10.5815/ijmecs.2019.11.04.
- [9] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter Sentiment Analysis on

- Worldwide COVID-19 Outbreaks,” *Kurdistan J. Appl. Res.*, pp. 54–65, 2020, doi: 10.24017/covid.8.
- [10] W. Y.-J. D. M. dan S. Informasi and undefined 2021, “Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier,” *Ejurnal.Teknokrat.Ac.Id*, vol. 2, no. 2, pp. 1–9, 2021, [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344>.
- [11] F. Syadid, “Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf- Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network,” *Skripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta*, pp. 1–89, 2019.
- [12] J. A. Septian, T. M. Fahrudin, and A. Nugroho, “Journal of Intelligent Systems and Computation 43,” *J. Intell. Syst. Comput.*, pp. 43–49, 2019, [Online]. Available: <https://t.co/9WloaWpfd5>.



KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Perbandingan Analisis Sentimen Terhadap 2 Jasa Ekspedisi Di Indonesia Dari Media Sosial Twitter Menggunakan Metode Naive Bayes Dan KNN (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, Labeling, 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 Analisis 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.