



***ANALISIS SENTIMEN UNTUK PENILAIAN PELAYANAN SITUS
BELANJA ONLINE MENGGUNAKAN ALGORITMA NAIVE BAYES***

TUGAS AKHIR



Lulu Anggraeni

41518110062

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022



***ANALISIS SENTIMEN UNTUK PENILAIAN PELAYANAN SITUS
BELANJA ONLINE MENGGUNAKAN ALGORITMA NAIVE BAYES***

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Lulu Anggraeni

41518110062

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518110062

Nama : Lulu Anggraeni

Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja
Online Menggunakan Algoritma Naïve Bayes

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.

Jakarta, 03 Agustus 2022


Lulu Anggraeni

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Lulu Anggraeni
NIM : 41518110062
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

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, 03 Agustus 2022

UNIVERSITAS
MERCU BUANA


Lulu Anggraeni

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Lulu Anggraeni
NIM : 41518110062
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

Menyatakan bahwa :


1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	
		Jurnal International Bereputasi	
	Disubmit/dipublikasikan di :	Nama Jurnal	: Lontar Komputer (Jurnal Ilmiah Teknologi Informasi)
		ISSN	: ISSN 2088-1541 (Print) ISSN 2541-5832 (Online)
		Link Jurnal	: https://ojs.unud.ac.id/index.php/lontar/index
		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


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

Jakarta, 03 Agustus 2022



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110062
Nama : Lulu Anggraeni
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

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, 03 Agustus 2022



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110062
Nama : Lulu Anggraeni
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

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, 03 Agustus 2022



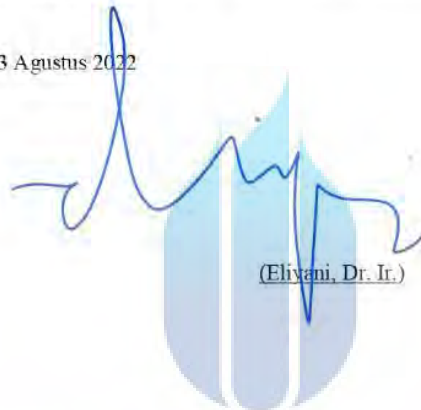
UNIVERSITAS
(Anis Cherid, SE, MTI)
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110062
Nama : Lulu Anggraeni
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs
Belanja Online Menggunakan Algoritma Naïve
Bayes

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, 03 Agustus 2022



(Elivani, Dr. Ir.)

UNIVERSITAS
MERCU BUANA

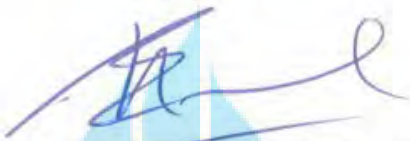
LEMBAR PENGESAHAN

NIM : 41518110062
Nama : Lulu Anggraeni
Judul Tugas Akhir : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

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, 03 Agustus 2022

Menyetujui,



Ir. Emil R. Kaburuan, Ph.D., IPM.
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT) (Ir. Emil R. Kaburuan, Ph.D., IPM.)
Koord. Tugas Akhir Teknik Informatika Ka. Prodi Teknik Informatika

ABSTRAK

Nama : Lulu Anggraeni
NIM : 41518110062
Pembimbing TA : Ir. Emil R. Kaburuan, Ph.D., IPM.
Judul : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

Belanja online merupakan situs belanja yang digemari oleh masyarakat pada saat ini dikarenakan menawarkan berbagai macam jenis barang dari yang diperlukan sampai tidak diperlukan, bukan hanya itu belanja online pun tidak perlu repot untuk dating kelokasi toko, dan pastinya transaksi jual beli pun mudah. Banyak masyarakat yang puas akan pelayanan yang dipersembahkan oleh situs belanja online yang masyarakat pilih dan banyak juga yang tidak puas. Pendapat yang diberikan masyarakat terhadap penilaian situs belanja online akan terungkap di media sosial, salah satunya adalah twitter. Penelitian ini akan melakukan analisis sentimen terhadap penilai opini pelanggan belanja online, yaitu Lazada, Bukalapak, Blibli dan Tokopedia pada twitter yang nantinya dapat digunakan untuk menentukan penilaian toko belanja online supaya masyarakat tidak salah dalam memilih situs toko belanja online. Dari postingan opini pelanggan yang di dapat akan dilakukan pelabelan sentiment yang menghasilkan 3 kelas berupa positive, negative dan netral setelah itu dilakukan pre-processing yaitu cleansing data, case folding, tokenizing, stop word removal dan stemming. Data yang sudah memiliki kelas data ini, akan di bagi menjadi data latih dan data uij dengan 3 skenario proses pengklasifikasi naïve bayes lalu data uji digunakan sebagai pengujian pada pengklasifikasian naïve bayes. Hasil klasifikasi dengan metode naïve bayes rata-rata menunjukkan akurasi sebesar 45,92%.

Kata kunci:

Situs belanja online, analisis sentiment, twitter, klasifikasi, Naive Bayes

ABSTRACT

Name : Lulu Anggraeni
Student Number : 41518110062
Counsellor : Ir. Emil R. Kaburuan, Ph.D., IPM.
Title : Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

Online shopping is a shopping site that is favored by the public at this time because it offers various types of goods from what is needed to not needed, not only that, online shopping does not need to bother going to the store location, and of course buying and selling transactions are easy. Many people are satisfied with the services offered by the online shopping sites that people choose and many are dissatisfied. The opinion given by the public on the assessment of online shopping sites will be revealed on social media, one of which is twitter. This study will conduct a sentiment analysis on the assessment of online shopping customer opinions, namely Lazada, Bukalapak, Blibli and Tokopedia on Twitter which can later be used to determine the assessment of online shopping stores so that people are not wrong in choosing an online shopping store site. From the post of customer opinion, sentiment labeling will be carried out which produces 3 classes in the form of positive, negative and neutral after that pre-processing is carried out, namely data cleansing, case folding, tokenizing, stop word removal and stemming. The data that already has this data class will be divided into training data and test data with 3 scenarios of the nave Bayes classifier process and then the test data is used as a test for the nave Bayes classification. The results of the classification using the nave Bayes method showed an average accuracy of 45.92%.

Key words:

Online Marketplace, Sentiment Analysis, Twitter, Classification, Naive Bayes

KATA PENGANTAR

Puji syukur kita panjatkan kepada ALLAH SWT yang telah melimpahkan Rahmat dan Karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes” dengan baik dan dalam jangka waktu yang telah ditentukan. Tugas akhir ini disusun untuk memenuhi syarat untuk LULUS sebagai sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, Tugas Akhir bisa saja tidak akan terselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktunya untuk memberikan bimbingan serta arahan dalam penyusunan Tugas Akhir ini.
2. Bapak Raka Yusuf, ST, MTI selaku dosen pembimbing akademik yang telah membimbing saya sejak awal mulai semester hingga sampai saat ini dan selalu memberikan motivasi serta arahan agar dapat lulus tepat waktu.
3. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM selaku Kepala Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.
4. Bapak Wawan Gunawan, S.Kom, MT selaku Koordinator Tugas Akhir Jurusan Teknik Informatika, Universitas Mercu Buana.
5. Keluarga dan orang tua yang selalu memberikan dukungan dan selalu mendoakan yang terbaik.
6. Teman-teman dan sahabat yang selalu memberikan dukungan dan memotivasi dalam melakukan penulisan tugas akhir ini agar dapat terselesaikan dengan baik.

Akhir kata, penulis berharap tugas akhir ini dapat menjadi bermanfaat bagi pembaca dan menambah wawasan pengetahuan semua pihak.

Jakarta, 03 Agustus 2022

Penulis

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	v
LEMBAR PENGESAHAN.....	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA	9
BAB 1. LITERATUR REVIEW.....	10
BAB 2. ANALISA DAN PERANCANGAN.....	12
BAB 3. SOURCE CODE	12
BAB 4. DATASET	20
BAB 5. TAHAPAN EKSPERIMEN.....	22
BAB 6. HASIL SEMUA EKSPERIMEN.....	28
DAFTAR PUSTAKA.....	30
LAMPIRAN DOKUMEN HAKI.....	33
LAMPIRAN KORESPONDENSI.....	34

NASKAH JURNAL

Analisa Sentimen Untuk Penilaian Pelayanan Situs BelanjaOnline Menggunakan Algoritma Naïve Bayes

Lulu Anggraeni^{a1}, Emil R.Kaburuan^{a2}
^aTeknik Informatika, Fakultas Ilmu Komputer
 Universitas Mercu Buana Jakarta
141518110062@student.mercubuana.ac.id*
[2emil.kaburuan@mercubuana.ac.id](mailto:emil.kaburuan@mercubuana.ac.id)

Abstrak

Belanja online merupakan situs belanja yang digemari oleh masyarakat pada saat ini dikarenakan menawarkan berbagai macam jenis barang dari yang diperlukan sampai tidak diperlukan, bukan hanya itu belanja online pun tidak perlu repot untuk dating kelokasi toko, dan pastinya transaksi jual beli pun mudah. Banyak masyarakat yang puas akan pelayanan yang dipersembahkan oleh situs belanja online yang masyarakat pilih dan banyak juga yang tidak puas. Pendapat yang diberikan masyarakat terhadap penilaian situs belanja online akan terungkap di media sosial, salah satunya adalah twitter. Penelitian ini akan melakukan analisis sentimen terhadap penilain opini pelanggan belanja online, yaitu Lazada, Bukalapak, Blibli dan Tokopedia pada twitter yang nantinya dapat digunakan untuk menentukan penilaian toko belanja online supaya masyarakat tidak salah dalam memilih situs toko belanja online. Dari postingan opini pelanggan yang di dapat akan dilakukan pelabelan sentiment yang menghasilkan 3 kelas berupa positive, negative dan netral setelah itu dilakukan pre-processing yaitu cleansing data, case folding, tokenizing, stop word removal dan stemming. Data yang sudah memiliki kelas data ini, akan di bagi menjadi data latih dan data uij dengan 3 skenario proses penglasifikasi naïve bayes lalu data uji digunakan sebagai pengujian pada penglasifikasian naïve bayes. Hasil klasifikasi dengan metode naïve bayes rata-rata menunjukkan akurasi sebesar 45,92%.

Kata kunci: Situs Belanja Online, Analisis Sentiment, Twitter, klasifikasi, Naive Bayes

1. Pendahuluan

Belanja online yaitu proses jual dan beli barang atau makanan oleh konsumen dari toko ke toko dengan teknologi yang canggih sebagai transaksi[1]. Dengan pesatnya zaman yang berkembang kini situs belanja *online* menjadi salah satu pilihan yang paling diminati masyarakat untuk berbelanja suatu produk karena dinilai memiliki banyak kelebihan seperti contohnya mereka tidak perlu datang langsung ke toko penjual, tidak adanya lagi pembatas untuk membeli sekalipun wilayah, dapat dilihat dan beli selama 24 jam, tawar menawar barang ditiadakan dikarenakan, banyak diskon, promo dan bahkan bebas biaya kirim. Akan tetapi dari kelebihan itu munculah beberapa kekurangan, yaitu penipuan tidak jarang terjadi pada saat melakukan pembelian, penjual tidak mengirimkan barang yang sudah dibeli, tidak sesuai barang yang ada digambar dengan aslinya, rusak barang serta jika terjadi kesalahan warna. Maka dari itu pembeli akan melampiaskan penilaian terhadap pelayanan yang diberikan oleh situs belanja onlen menggunakan sosial media tersebut.

Saat ini, beberapa masyarakat menggunakan media sosial untuk memberikan pendapat mereka terhadap sesuatu yang terjadi pada dirinya atau dunia. Terdapat banyak sosial media terkenal seperti Twitter, Facebook, Instagram dll. Salah satu dari banyaknya media sosial yang terkenal dan sering digunakan dalam menyampaikan pendapat adalah twitter. Berbagai pendapat pelanggan perihal situs belanja online yang ada di Twitter, pelanggan tidak dapat menyimpulkan situs belanja online mana yang mendapatkan penilaian paling baik. Hal ini dapat menyebabkan banyaknya komentar di twitter bahkan

trending topic yang ada di twitter pun hanya menampilkan topik yang sedang banyak diperbincangkan tanpa memberikan suatu kesimpulan[2].

Analisis sentimen atau biasa disebut juga dengan *opinion mining* adalah sebuah proses mendapatkan opini pengguna terhadap beberapa topik atau teks yang disampaikan pengguna[3]. Dalam pengertian lain, sebuah proses menentukan apakah sepotong kalimat itu bernilai positif, negatif atau netral. Penelitian ini membahas tentang bagaimana analisis sentimen terhadap data opini pelanggan *belanja online* pada Twitter yang digunakan untuk menghasilkan data opini tersebut masuk opini positif atau negatif, sehingga nantinya bisa dimanfaatkan untuk menentukan situs belanja online yang dikategorikan baik.

Beberapa penelitian berkaitan dengan topik analisis sentimen sudah banyak dilakukan, diantaranya : penelitian *An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis*[5] menggunakan enam metode untuk klasifikasi yaitu Lexicon-based classifier, NB, Bayesian Network, SVM (Support Vector Machine), C4.5 (Decision Tree), Random Forest serta satu metode yang disebut dengan Ensemble Classifier yang menggabungkan lima metode (NB, Bayesian Network, SVM, C4.5 dan Random Forest) untuk mendapatkan akurasi yang lebih tinggi. Penelitian ini menggunakan empat kelas yaitu kelas positif (4288 tweet), negatif (35876 tweet), netral (40987 tweet) dan *irrelevant* (26715 tweet). Perolehan akurasi masing-masing saat tidak dikombinasikan dengan dataset dua kelas (menghilangkan kelas netral dan irrelevant) adalah Lexicon Based 67.9%, Naïve Bayesian 90%, Bayesian Network 91.4%, SVM 84.6%, Random Forest 89.8%. Metode Lexicon Based tidak ikut dalam kombinasi karena perolehan akurasi paling sedikit yaitu 67,9%, perolehan akurasi ensemble dengan dataset dua kelas yaitu 91.7% sedangkan perolehan akurasi ensemble untuk dataset tiga kelas yaitu 84.2%.

Dalam *Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier* [6] menggunakan dua metode supervised dengan dua dataset yaitu film dan hotel, semakin banyak data latih yang diinputkan semakin bagus akurasi yang diperoleh pada algoritma NB dengan dataset film tetapi untuk metode K-NN akurasi diperoleh secara random.

Penelitian calon presiden, diteliti opini masyarakat terhadap calon presiden Indonesia tahun 2014[4] yaitu pasangan Prabowo Subianto-Hatta Rajasa dan Joko Widodo-Jusuf Kalla. Penelitian [4] menggunakan NB untuk pengklasifikasian dokumen, selain sederhana metode ini mempunyai akurasi yang tinggi jika diaplikasikan dalam data yang besar dan beragam. Data dalam penelitian ini diambil dalam tiga periode yaitu sebelum pemilu legislatif, saat diadakan pemilu legislatif dan setelah deklarasi pengumuman pemilu legislatif kemudian dari data tersebut penulis mengelompokkan opini masyarakat apakah bersifat positif, negatif atau netral. Hasil dari polaritas sentimen, Prabowo Subianto-Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 26,4% untuk sentimen negatif dan 25,9% untuk sentimen netral sedangkan Joko Widodo-Jusuf Kalla mendapatkan 37,6% sentimen positif, 34,4% sentimen negatif dan 27,9% sentimen netral dengan tingkat akurasi sebesar 90%.

Penelitian Klasifikasi Teks dengan NB untuk Pengelompokkan Teks Berita dan Abstrak Akademis[7]. Melakukan sebanyak 7 kali eksperimen untuk dokumen berita maupun dokumen abstrak akademis pada eksperimen pertama dengan perbandingan data latih dan data uji 9:1 dari dokumen seluruhnya mendapatkan hasil akurasi yang tertinggi lalu data latih akan dikurangi jumlahnya akurasi berkurang juga. Menggunakan data latih sebesar 50% dari data semuanya memperoleh hasil akurasi lebih dari 75%.

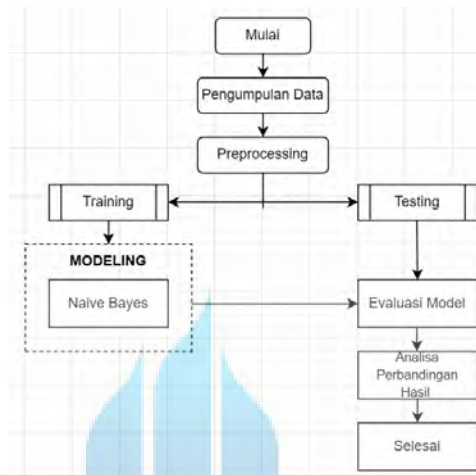
Penelitian Analisis Opini Terhadap Fitur *Smartphone* pada Ulasan *Website* Berbahasa Indonesia[8]. Data dikumpulkan dengan cara web scraping merupakan pengambilan data penilaian dari website target. Target yang masuk ke dalam analisa adalah BOLT 4G Powerphone IVO, Samsung Galaxy Grand I9082 White, Apple iPhone 5S 16GB Gold White, Microsoft Lumia Orange dan Nokia Lumia 520 Black. Sebelum dilakukannya analisa data yang sudah dilakukan pengumpulan dari web harus ketahap preprocessing agar mendapatkan penilaian data bersih lalu Lexicon based, rule bahasa dan score function digunakan dalam menganalisis opini. Hasil pengujian yang diperoleh

rata-rata nilai recall dan precision sebesar 0.63 dan 0.72 sedangkan akurasi sebesar 81.76%.

Pada penelitian ini, dalam melakukan analisis sentimen, menggunakan algoritma Naïve Bayes. Algoritma yang paling umum dan dapat mudah mengerti dan digunakan untuk proses klasifikasi yang bersumber dari data twitter yang belum memiliki pelabelan kelas data untuk pengklasifikasian kedalam kelas opini positif, negatif dan netral.

2. Metode Penelitian

Pada penelitian ini terdapat beberapa tahapan eksperimen yang dilakukan. Untuk secara keseluruhan semua tahapan dapat dilihat pada gambar 1 dibawah dan dengan penjelasannya



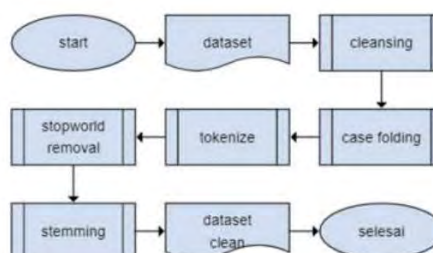
Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pengumpulan data yang berasal dari *tweet* maupun *re-tweet* masyarakat di *twitter*. Data diambil hanya dari situs *twitter.com* di Indonesia yaitu Lazada, Bukalapak, Blibli dan Tokopedia. Data diperoleh dari proses *crawling* yaitu proses pengambilan kalimat opini dari *twitter* melalui API *twitter* dengan rentan waktu dari tanggal 1 juli 2022 hingga 6 juli 2022, sehingga data yang berhasil dikumpulkan yaitu sebanyak 1198 data ulasan. Kemudian data tersebut diseleksi dan *cleaning* kembali dengan menghapus beberapa ulasan yang tidak termasuk dalam ulasan penilaian pelayanan situs belanja, sehingga data menjadi 300 data.

2.2 Pra Pengolahan (*Pre-Processing*) Data

Data *tweet* yang telah diambil dari twitter masih berupa data mentah maka dari itu dilakukan tahap *preprocessing* untuk mendapatkan data bersih agar dapat diproses ke tahap selanjutnya. Tahapan yang dilakukan adalah :



Gambar 2. Pre-Processing

- 1) *Cleansing* : Data tweet akan dilakukan pengurangan noise. Pengurangan yang dimaksud berupa kata yang di hilangkan bersifat tidak penting, yaitu URL, hashtag (#), username (@), email, emoticon (:@, :*, :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya. Contoh proses *cleansing* data terdapat pada Tabel 1.

Tabel 1. Cleansing Data

Sebelum	Sesudah
@bukalapak Halo..halo Min..DM saya tolong di FU dong. Kenapa ongkos kirim transaksi penjualan toko saya dibebankan ke toko saya ya? Detail nomor transaksinya sudah lama saya DM. Terima kasih	Halo halo Min DM saya tolong di FU dong Kenapa ongkos kirim transaksi penjualan toko saya dibebankan ke toko saya ya Detail nomor transaksinya sudah lama saya DM Terima kasih

- 2) *Case Folding* : Pada data *tweet*, akan terdapat beberapa bentuk huruf yang berbeda, pada tahap case folding ini dimana bentuk huruf akan berubah menjadi huruf kecil (*lower case*) dengan kata lain penyesuaian huruf. Contoh proses *case folding* data terdapat pada Tabel 2.

Tabel 2. Case Folding Data

Sebelum	Sesudah
Halo halo Min DM saya tolong di FU dong Kenapa ongkos kirim transaksi penjualan toko saya dibebankan ke toko saya ya Detail nomor transaksinya sudah lama saya DM Terima kasih	halo halo min dm saya tolong di fu dong kenapa ongkos kirim transaksi penjualan toko saya dibebankan ke toko saya ya detail nomor transaksinya sudah lama saya dm terima kasih

- 3) *Tokenizing* : *tokenizing* (parsing) merupakan string yang terpotong pada tiap kata yang tersusun, dengan kata lain memenggal beberapa kata dari sebuah kalimat. Contoh proses *Tokenizing* data terdapat pada Tabel 3.

Tabel 3. Tokenizing

Sebelum	Sesudah
halo halo min dm saya tolong di fu dong Kenapa ongkos kirim transaksi penjualan toko saya dibebankan ke toko saya ya detail nomor transaksinya sudah lama saya dm terima kasih	"halo", "halo", "min", "dm", "saya", "tolong", "di", "fu", "dong", "kenapa", "ongkos", "kirim", "transaksi", "penjualan", "toko", "saya", "dibebankan", "ke", "toko", "saya", "ya", "detail", "nomor", "transaksinya", "sudah", "lama", "saya", "dm", "terima", "kasih"

- 4) *Stopword Removal* : memproses hilangnya sebuah kata dari daftar yang tidak sesuai pada sebuah kalimat yang semestinya dihilangkan seperti "yang", "di", "ke", "itu" dan lain sebagainya. Contoh proses *stopword removal* data terdapat pada Tabel 4.

Tabel 4. Stopword Removal

Sebelum	Sesudah
"halo", "halo", "min", "dm", "saya", "tolong", "di", "fu", "dong", "Kenapa", "ongkos", "kirim", "transaksi", "penjualan", "toko", "saya", "dibebankan", "ke", "toko", "saya", "ya", "detail", "nomor", "transaksinya", "sudah", "lama", "saya", "dm", "terima", "kasih"	halo min dm saya tolong fu kenapa ongkos kirim transaksi penjualan toko saya dibebankan toko saya detail nomor transaksinya sudah lama saya dm terima kasih

- 5) *Stemming* : membuat sebuah kata yang berhimpunan menjadi kata dasar yang sesuai dengan standar Bahasa Indonesia. Contoh proses *stemming* data terdapat pada Tabel 5.

Tabel 5. Stemming

Sebelum	Sesudah
Halo min dm saya tolong fu kenapa ongkos kirim transaksi penjualan toko saya dibebankan toko saya detail nomor transaksinya sudah lama saya dm terima kasih	Halo min dm saya tolong fu kenapa ongkos kirim transaksi jual toko saya beban toko saya detail nomor transaksi sudah lama saya dm terima kasih

2.4 Pembagian Data

Data yang telah dilakukan tahapan *text preprocessing* selanjutnya data dibagi menjadi data *training* dan data *testing*. Pembagian data dilakukan dengan rasio 90% data latih dan 10% data uji.

2.5 Pembuatan Model

Metode Naïve Bayes merupakan metode pengklasifikasian dalam memproses teks yang diperlukan pada sebuah analisis sentimen. Metode ini memiliki potensi yang baik dalam sebuah klasifikasi hal presisi dan komputasi data. [25]. Naïve Bayes sudah biasa digunakan terkait klasifikasi terutama pada sumber data yang berasal dari Twitter[26]. Memperhitungkan nilai dari probabilitas dalam Naïve Bayes menggunakan pendekatan algoritma Bayes. Berikut adalah rumus dari naïve bayes:

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

Dimana:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

2.5 Evaluasi Model

Setelah data di *training* menggunakan algoritma naïve bayes, selanjutnya dilakukan evaluasi terhadap model. Evaluasi akan diukur menggunakan *confusion matrix* untuk mendapatkan nilai akurasi.

Accuracy =

$$\frac{(TPos+TNeg+TNet)}{(TPos+FPosNeg+FPosNet+FNegPos+TNeg+FNegNet+FNetPos+FNetNeg+TNet)}$$

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	TPos	FPosNeg	FPosNet
	Negatif	FNegPos	TNeg	FNegNet
	Netral	FNetPos	FNetNeg	TNet

Gambar 3. Confusion Matrix

3. Hasil dan Pembahasan

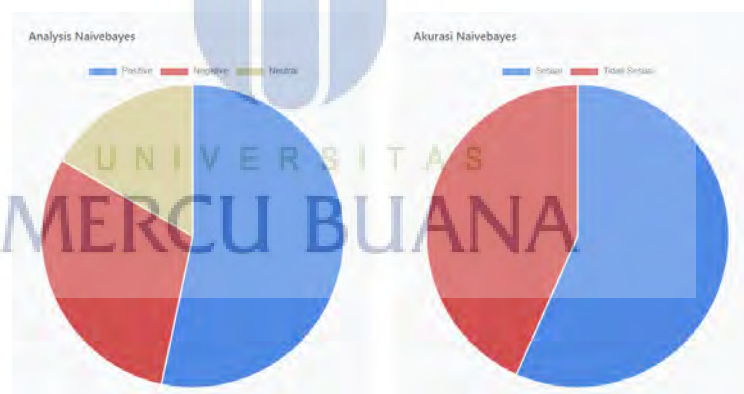
Eksperimen dilakukan dengan melakukan beberapa percobaan. Percobaan dilakukan dengan mengubah komposisi pembagian data uji dan data latih.

3.1 Pembagian data latih dan data uji

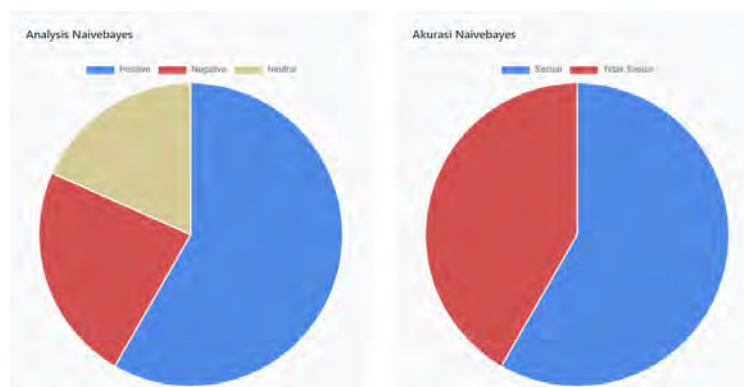
Eksperimen pembagian data latih dan uji akan dilakukan 3 tahap skenario, yaitu pertama dengan membagi data pada komposisi 90% data latih dan 10% data uji, percobaan kedua dengan komposisi 80% data latih dan 20% data uji, dan percobaan ketiga dengan komposisi 70% data latih dan 30% data uji. Pada tabel 6 menyajikan seluruh hasil pengujian dengan 3 skenario pembagian data latih dan data uji.

Tabel 6. Hasil pengujian pembagian data latih dan data uji

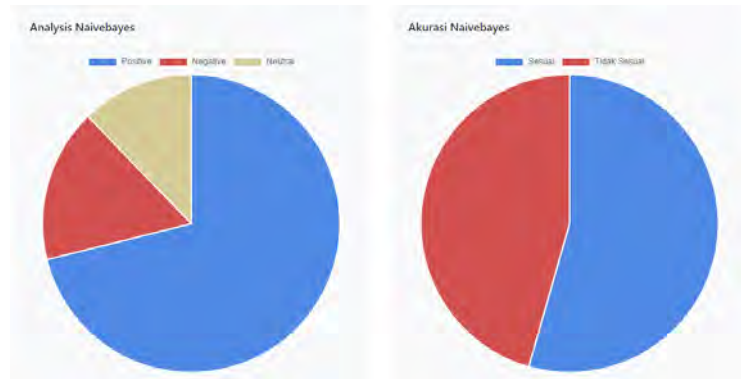
Naïve Bayes	
komposisi	Accuracy
10/90	43,33%
20/80	53,33%
30/70	41,11%
TextBlob	
komposisi	Accuracy
10/90	10%
20/80	21,66%
30/70	33,33%



Gambar 4. Diagram Lingkaran Naïve Bayes dengan komposisi 10/90



Gambar 5. Diagram Lingkaran Naïve Bayes dengan komposisi 20/80



Gambar 6. Diagram Lingkaran Naïve Bayes dengan komposisi 30/70

4. Simpulan

Berdasarkan hasil dan pembahasan diatas, kesimpulan yang didapat adalah analisis sentimen terhadap situs belanja online menggunakan Algoritma Naïve Bayes, yang dilakukan pada dua tahap, yaitu tahap data latih dan tahap data uji dengan dilakukan 3 tahap skenario. Hasil dari semua komposisi pembagian data yang telah dilakukan menunjukan tingkat akurasi yang baik pada ketiga skenario. Untuk algoritma Naïve Bayes memiliki tingkat akurasi tertinggi pada komposisi 80% data latih dan 20% data uji dengan tingkat akurasinya adalah 53.33%. Dan Hasil klasifikasi dengan metode naïve bayes rata-rata menunjukkan akurasi sebesar 45,92%.

Daftar Pustaka

- [1] A. Ishak, "Analisis Kepuasan Pelanggan Dalam Belanja Online : Sebuah Studi Tentang Penyebab (Antecedents) dan Konsekuensi (Consequents)," *Jurnal Siasat Bisnis*, Vol. 16 No. 2, , Hal : 141-154, Juli 2012.
- [2] D. G. Nugroho, Y. H. Chrishanto, and A. Wahana, "Analisis Sentimen pada Jasa Ojek Online Menggunakan Metode Naïve Bayes," *Jur. Inform. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Jenderal Achmad Yani*, pp. 156–161, 2015.
- [3] B. S. D and P. D. Gore, "Sentiment Analysis on Twitter Data Using Support Vector Machine," *International Journal of Computer Science Trends and Technology (IJCSST) – Volume 4 Issue 3*, pp. 365–370, May - Jun 2016.
- [4] F. Nurhuda and S. W. Sihwi, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *JURNAL ITSMART Vol 2. No 2. ISSN : 2301–7201*, Desember 2013
- [5] Y. Wan, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis," *IEEE 15th International Conference on Data Mining Workshops*, 2015.
- [6] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016.
- [7] A. Hamzah, "Sentiment Analisis Untuk Memanfaatkan Saran Kuesioner dalam Evaluasi Pembelajaran dengan Menggunakan Naive Bayes Classifier (NBC)," *Pros. Semin. Nas. Apl. Sains Teknol.*, pp. 17–24, November, 2014.
- [8] D. Setyawan and E. Winarko, "Analisis Opini Terhadap Fitur Smartphone Pada Ulasan Website Berbahasa Indonesia," *IJCCS (Indonesian J. Comput. Cybern.*

- Syst., vol. 10, no. 2, p. 183, 2016.
- [9] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier," IEEE International Conference on Big Data, pp. 99–104, 2013.
 - [10] A. P. Wijaya and H. A. Santoso, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," Journal of Applied Intelligent System, Vol.1, No. 1, pp. 48–55, Februari 2016.
 - [11] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in Proceedings of the 10th European Conference on Machine Learning, 1997.
 - [12] I. H. Witten, E. Frank, M. A. Hall, "Data mining: practical machine learning tools and techniques. 3rd Edition," San Francisco , MA: Morgan Kaufmann, 2011.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “ANALISIS SENTIMEN UNTUK PENILAIAN PELAYANAN SITUS BELANJA ONLINE MENGGUNAKAN ALGORITMA NAIVE BAYES”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review , Analisa perancangan , dataset yang digunakan , tahapan eksperimen , dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan Analisa perancangan dan alur sistem rekomendasi. Bagian III menjelaskan mengenai source code yang digunakan pada penelitian ini. Bagian VI menjelaskan mengenai dataset yang digunakan. Bagian V memuat tahapan eksperimen yang disajikan dalam gambar dan table dengan penjelasan dari setiap tahapan. Bagian VI merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.