

SKRIPSI

**ANALISIS SENTIMEN TERHADAP *REVIEW* PRODUK E-
MARKETPLACE MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***



Disusun Oleh :

ATANG SUHARNA

2015230108

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS DARMA PERSADA
JAKARTA
2021**

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Atang Suharna

NIM : 2015230108

Fakultas : Teknik

Jurusan : Teknologi Informasi

Menyatakan bahwa Laporan Tugas Akhir ini saya susun sendiri berdasarkan hasil peninjauan, penelitian lapangan, wawancara serta memadukannya dengan buku-buku, literature atau bahan-bahan referensi lain yang terkait dan relevan di dalam penyelesaian Laporan Kerja Praktek ini.

Demikian pernyataan ini penulis buat dengan sesungguhnya

Jakarta, 10 Juli 2021



Atang Suharna

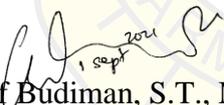
LEMBAR PENGESAHAN

ANALISIS SENTIMEN TERHADAP REVIEW PRODUK E-MARKETPLACE
MENGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST
NEIGHBOR

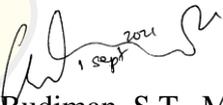
Disusun oleh :

Nama : Atang Suharna

NIM : 2015230108


Adam Arif Budiman, S.T., M. Kom.

Pembimbing Laporan


Adam Arif Budiman, S.T., M. Kom.

Kajur Teknologi Informasi

LEMBAR PENGUJI SKRIPSI

Laporan SKRIPSI yang berjudul :

“ANALISIS SENTIMEN TERHADAP REVIEW PRODUK
E-MARKETPLACE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN
K-NEAREST NEIGHBOR” ini telah ujikan pada tanggal

19 Agustus 2021



Penguji I

A handwritten signature in black ink, appearing to read 'Bagus'.

Bagus Tri Mahardhika, MMSI

Penguji 2

A handwritten signature in black ink, appearing to read 'Andi'.

Andi Susilo, M.Kom

Penguji 3

A handwritten signature in black ink, appearing to read 'Herianto'.

Herianto, MT

LOGBOOK BIMBINGAN SKRIPSI

NAMA MAHASISWA : Atang Suharna
NIM : 2015230108
DOSEN PEMBIMBING : Adam Arif Budiman, ST.,M.KOM
JUDUL KERJA PRAKTEK : ANALISIS SENTIMEN TERHADAP REVIEW
PRODUK MENGGUNAKAN ALGORITMA
NAÏVE BAYES DAN K-NEAREST NEIGHBOR
(STUDI KASUS: TOKO MIRING)
SEMESTER / TA : Ganjil, 2020/2021



**JURUSAN TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS DARMA PERSADA**



LEMBAR BIMBINGAN SKRIPSI
TEKNOLOGI INFORMASI – DARMA PERSADA

NIM : 2015230108

NAMA LENGKAP : Atang Suharna

DOSEN PEMBIMBING : Adam Arif Budiman, ST.,M.Kom

JUDUL : ANALISIS SENTIMEN TERHADAP REVIEW
PRODUK MENGGUNAKAN ALGORITMA
NAÏVE BAYES DAN K-NEAREST NEIGHBOR
(STUDI KASUS: TOKO MIRING)

No	Tanggal	Materi	Paraf Dosen Pembimbing
1	26 Mei 2021	Penyerahan BAB I	
2	27 Mei 2021	Penyerahan BAB II	
3	27 Mei 2021	Penyerahan BAB III	
4	21 Juni 2021	Perbaikan BAB III	
5	5 Juli 2021	Perbaikan Penulisan BAB III	
6	6 Juli 2021	Penyerahan BAB IV	
7	6 Juli 2021	Penyerahan BAB V	
8	10 Juli 2021	Demo Aplikasi	
9	16 Juli 2021	Laporan Siap Jilid	
10	16 Juli 2021	Jurnal	
11		
12			Acc seminar isi ····

Jakarta, 16 Juli 2021

Dosen Pembimbing

Adam Arif Budiman, ST.,M.Kom

FORM PERSETUJUAN

SEMINAR SKRIPSI

Dosen pembimbing dari Mahasiswa berikut

Nama : Atang Suharna
NIM : 2015230108
Pembimbing : Adam Arif Budiman, ST.,M.Kom
Judul : ANALILSIS SENTIMEN TERHADAP REVIEW
PRODUK MENGGUNAKAN ALGORITMA
NAÏVE BAYES DAN K-NEAREST NEIGHBOR
(STUDI KASUS: TOKO MIRING)
Semester /TA : Ganjil, 2020/2021

Menyetujui untuk mengikuti seminar Skripsi, dimana mahasiswa tersebut telah melengkapi:

No	Kelengkapan	Paraf pembimbing	Keterangan
1	Laporan Skripsi dijilid 3 rangkap yang telah di setujui pembimbing		
2	Power Point presentasi		
3	Program Aplikasi TA telah mencapai 100%		
4	Artikel Jurnal		

Jakarta, 16 Juli 2021
Mengetahui,
Dosen Pembimbing

Adam Arif Budiman, ST.,M.Kom



UNIVERSITAS DARMA PERSADA

Jl. Radin Inten II (Terusan Casablanca) Pondok Kelapa – Jakarta 13450

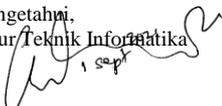
Telp. (021) 8649051, 8649053, 8649057 Fax. (021) 8649052

E-mail : humas@unsada.ac.id Home page : <http://www.unsada.ac.id>

LEMBAR PERBAIKAN SIDANG SKRIPSI

Nama : ATANG SUHARNA
Nim : 2015230108
PRODI : Teknologi Informasi
Dosen Pembimbing : Adam Arif Budiman, M.Kom
Judul : ANALISIS SENTIMEN TERHADAP REVIEW PRODUK ONLINE SHOP
MENGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR

No.	Keterangan	Dosen
1.	Penerapan metode pada perancangan sistem	Bagus Mahardika 
2.	1. Judul diperbaiki: online shop ganti dengan e-marketplace 2. Rumusan masalah spesifik dengan kendala-kendala bagi e-marketplace (Lazada), munculkan kondisi-kondisi nyata yang terjadi 3. Kesimpulan dan saran harus menjawab masalah dan kesesuaian dari tujuan dibangunnya analisis sentimen ini.	Andi Susilo 
3.		Herianto 05032021 

Mengetahui,
Kajur Teknik Informatika


Bila form ini tidak mencukupi silakan ditulis di balik lembar ini →

SKRIPSI

**ANALISIS SENTIMEN TERHADAP *REVIEW* PRODUK E-
MARKETPLACE MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***



Disusun Oleh :

ATANG SUHARNA

2015230108

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS DARMA PERSADA

JAKARTA

2021

KATA PENGANTAR

Puji syukur saya panjatkan kehadirat Tuhan Yang Maha Esa yang telah memberikan karunia-Nya kepada saya sehingga dapat menyelesaikan penulisan laporan tugas akhir. Dalam pembuatan laporan dan aplikasi tugas akhir ini, tidak jarang saya menemui berbagai macam kesulitan dan hambatan. Untuk itu pada kesempatan kali ini saya ingin mengucapkan terima kasih kepada pihak-pihak yang secara langsung maupun tidak langsung membantu dalam pembuatan aplikasi dan penulisan laporan ini. Saya menyampaikan ucapan terima kasih secara khusus kepada :

- 1 Dekan Fakultas Teknik Universitas Darma Persada Ir. Agus SunSugiharto,MT.
- 2 Ketua Jurusan Teknik Informatika Adam Arif Budiman, ST., M.Kom.
- 3 Dosen Pembimbing Adam Arif Budiman, ST., M.Kom yang telah sabar membimbing saya dalam penyusunan Laporan Tugas Akhir.
- 4 Dosen-dosen Teknik Informatika Universitas Darma Persada yang telah memberikan ilmu yang sangat bermanfaat kepada saya.
- 5 Keluarga saya tercinta terutama orang tua saya yang selalu memberikan doa dan semangat kepada saya.

- 6 Rekan - rekan TIF 2015 yang memberikan semangat kepada saya. Saya menyadari bahwa pembuatan laporan dan aplikasi tugas akhir ini masih jauh dari kata sempurna karena keterbatasan pengetahuan, kemampuan dan pengalaman yang dimiliki. Oleh karena itu, kritik dan saran dari

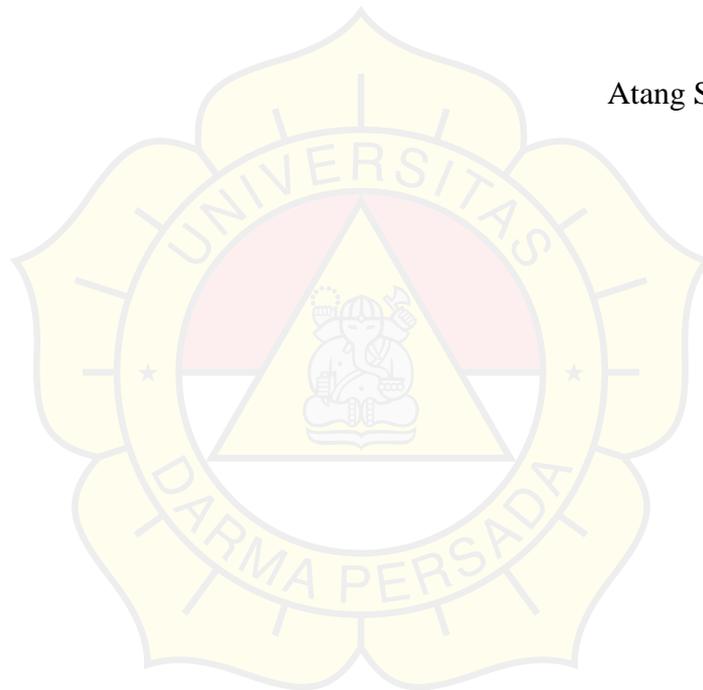


semua pihak yang membangun selalu saya harapkan demi perbaikan yang lebih baik dikemudian hari.

Akhir kata semoga penulisan laporan tugas akhir ini dapat memberikan manfaat bagi penulis sendiri maupun para pembacanya.

Jakarta, 10 Juli 2021

Atang Suharna



ABSTRAK

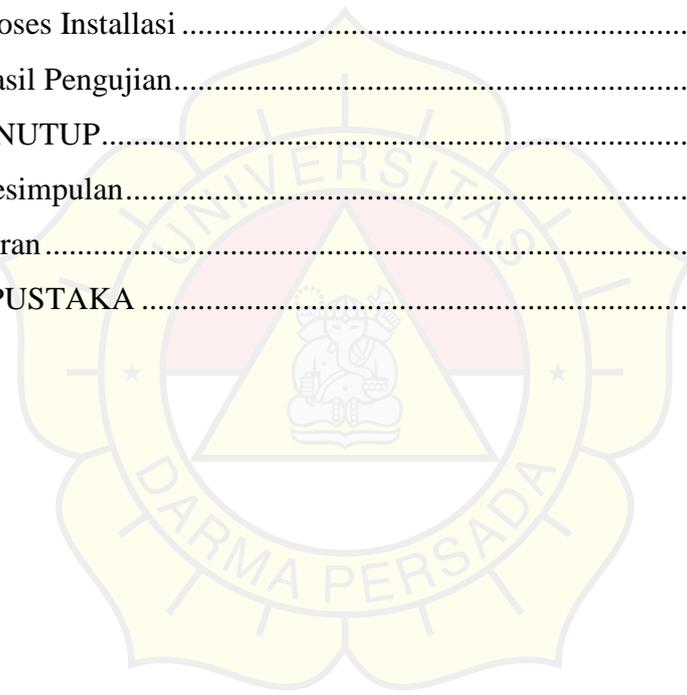
Tren Sentimen merupakan bidang studi yang menganalisa opini masyarakat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap sebuah produk, pelayanan, organisasi dan perhimpunan, seorang tokoh, dan isu atau masalah serta peristiwa yang terjadi pada masyarakat itu sendiri. pada penelitian kali ini, peneliti analisis sentimen terhadap *review* produk menggunakan algoritma naïve bayes dan k-nearest neighbor dengan studi kasusnya adalah *E-Marketplace*. Dari hasil yang didapatkan dari kedua metode peneliti melakukan perbandingan untuk mendapatkan nilai akurasi yang paling baik dari kedua metode tersebut. Metode yang akan digunakan dalam penelitian ini dengan cara wawancara, kuesioner, observasi, studi pustaka untuk memperoleh data yang akan dimasukkan ke sistem. Pengembangan sistem yang digunakan adalah *System Development Life Cycle* (SDLC) dengan cara menspesifikasikan yang dibutuhkan oleh *User*, membuat perancangan sistem menggunakan *Unified Modelling Language* (UML), mengimplementasikan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) kedalam bahasa pemrograman Python dan MySQL sebagai databasenya. Alur metodologi algoritma *K-Nearest Neighbor* (KNN) yaitu Input dataset, preprocessing, TF-IDF, Normalisasi TF-IDF, Cosine Similarity, Hitung Distance, Voting dan *Confusion Matrix*. Sedangkan Alur metodologi algoritma *Naïve Bayes* yaitu Input dataset, preprocessing, *Term Frequency*, *Prior Probability*, *conditional Probability*, Hitung Posterior, dan *Confusion Matrix*. Sistem yang telah dibangun akan diuji menggunakan metode *Blackbox Testing* dan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi dari kedua metode tersebut. *Text mining* adalah salah satu penambangan informasi yang berguna dari data – data yang berupa tulisan dokumen atau text dalam bentuk klasifikasi maupun clustering. *Preprocessing* merupakan tahap untuk memproses data teks untuk didapat dianalisa. Pada penelitian ini mendapatkan hasil akurasi dari setiap algoritma yang digunakan, seperti hasil akurasi yang didapatkan menggunakan KNN adalah 78,26% dan lebih baik dari pada hasil akurasi yang didapatkan menggunakan Naive Bayes 73,91%.

Kata kunci: *Analisa Sentimen, naïve bayes dan k-nearest neighbor, SDLC, UML, Python dan MySQL*

DAFTAR ISI

LEMBAR PENGESAHAN	ii
KATA PENGANTAR	iii
ABSTRAK	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi	4
1.6.1. Metode Penelitian.....	4
1.6.2. Metodologi Pengembangan Sistem.....	5
1.7 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	8
2.1 Text Mining	8
2.2 Preprocessing.....	8
2.3 Naïve Bayes Classifier	9
2.4 K Nearest Neighbor.....	10
2.5 Confusion Matrix	11
2.6 Jurnal Penelitian Yang Berkaitan	12
BAB III DESAIN DAN PERANCANGAN SISTEM.....	14
3.1 Analisa Permasalahan.....	14
3.2 Analisa Kebutuhan	14
3.2.1 Analisa Kebutuhan Fungsional.....	14
3.1.1. Analisa Kebutuhan Non-Fungsional	15
3.3 Perancangan Sistem.....	15
3.3.1 Flowchart Sistem Usulan.....	15

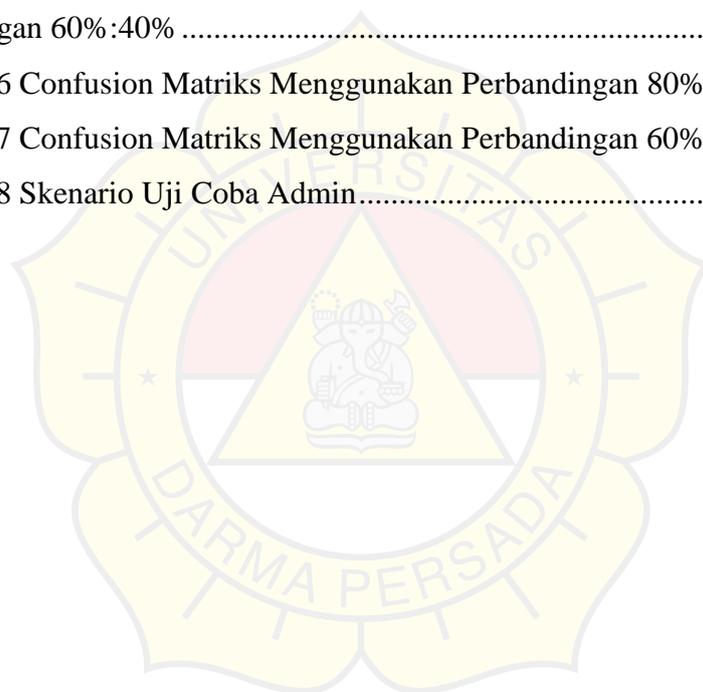
3.3.2	Unified Modeling Language (UML)	17
3.3.3	Rancangan Database	23
3.3.4	Rancangan Interface	24
BAB IV	IMPLEMENTASI SISTEM	28
4.1	Hasil Implementasi Sistem	28
4.1.1	Admin	28
4.1.2	Pengguna.....	31
4.2	Hasil Implementasi Perbandingan Algoritma KNN dan Naïve Bayes... 34	
4.2.1	Implementasi Algoritma KNN.....	34
4.2.2	Imlementasi Algoritma Naïve Bayes	48
4.3	Proses Instalasi	58
4.4	Hasil Pengujian.....	60
BAB V	PENUTUP.....	62
5.1	Kesimpulan.....	62
5.2	Saran	62
DAFTAR PUSTAKA	64



DAFTAR TABEL

Tabel 3. 1 Rancangan table user	23
Tabel 3. 2 Rancangan tabel data review.....	23
Tabel 4. 2.1 Confusion Matrix	10
Tabel 4. 2 Dataset Feedback Pelanggan.....	34
Tabel 4. 3 Hasil Preprocessing.....	35
Tabel 4. 4 Hasil Penentuan Kelas Berdasarkan Rating.....	37
Tabel 4. 5 Hasil Perhitungan Term Untuk Setiap Dokumen	37
Tabel 4. 6 Hasil Pembagian Data Training	38
Tabel 4. 7 Hasil Pembagian Data Testing.....	38
Tabel 4. 8 Perhitungan df dan idf Menggunakan Data Training.....	38
Tabel 4. 9 Perhitungan wtf Menggunakan Data Training.....	39
Tabel 4. 10 Perhitungan wtf Menggunakan Data Testing.....	39
Tabel 4. 11 Perhitungan wtd Menggunakan Data Training	40
Tabel 4. 12 Perhitungan wtd Menggunakan Data Testing.....	40
Tabel 4. 13 Perhitungan Normalisasi wtd Menggunakan Data Training.....	41
Tabel 4. 14 Perhitungan Normalisasi wtd Menggunakan Data Testing	41
Tabel 4. 15 Hasil Perhitungan Cosine Similarity.....	42
Tabel 4. 16 Hasil Perhitungan Cosine Similarity	42
Tabel 4. 17 Hasil Pengurutan Jarak Terdekat Dengan Data Uji	43
Tabel 4. 18 Hasil Pengurutan Jarak Terdekat Dengan Data Uji	43
Tabel 4. 19 Hasil Pengambilan Dokumen Terdekat Sebanyak k.....	43
Tabel 4. 20 Hasil Pengambilan Dokumen Terdekat Sebanyak k.....	44
Tabel 4. 21 Hasil Pengambilan Dokumen Sebanyak k	44
Tabel 4. 22 Perbandingan Kelas Prediksi dengan Kelas Aktual Menggunakan Perbandingan 80%:20%	45
Tabel 4. 23 Perbandingan Kelas Prediksi dengan Kelas Aktual Menggunakan Perbandingan 60%:40%	45
Tabel 4. 24 Confusion Matriks Menggunakan Perbandingan 80%:20%	45
Tabel 4. 25 Confusion Matriks Menggunakan Perbandingan 60%:40%	45
Tabel 4. 26 Dataset Feedback Pelanggan.....	48

Tabel 4. 27 Hasil Preprocessing.....	49
Tabel 4. 28 Hasil Penentuan Kelas Berdasarkan Rating.....	50
Tabel 4. 29 Hasil Perhitungan TF	51
Tabel 4. 30 Hasil Pembagian Data Training	51
Tabel 4. 31 Hasil Pembagian Data Testing.....	51
Tabel 4. 32 Hasil Perhitungan Multinomial Model	53
Tabel 4. 33 Hasil Perhitungan Posterior	55
Tabel 4. 34 Perbandingan Kelas Prediksi dengan Kelas Aktual Menggunakan Perbandingan 80%:20%	55
Tabel 4. 35 Perbandingan Kelas Prediksi dengan Kelas Aktual Menggunakan Perbandingan 60%:40%	56
Tabel 4. 36 Confusion Matriks Menggunakan Perbandingan 80%:20%	56
Tabel 4. 37 Confusion Matriks Menggunakan Perbandingan 60%:40%	56
Tabel 4. 38 Skenario Uji Coba Admin.....	57



DAFTAR GAMBAR

Gambar 3. 1 Flowchart.....	16
Gambar 3. 2 Usecase Diagram.....	17
Gambar 3. 3 Activity Diagram Login	18
Gambar 3. 4 Activity Diagram Kelola Data Pengguna.....	19
Gambar 3. 5 Activity Diagram Kelola Data Review	19
Gambar 3. 6 Activity Diagram Kelola Hasil Analisa	20
Gambar 3. 7 Sequence Diagram login	21
Gambar 3. 8 Sequence Diagram Kelola Data Pengguna.....	21
Gambar 3. 9 Sequence Diagram Kelola Data review	22
Gambar 3. 10 Sequence Diagram Hasil Analisa.....	22
Gambar 3. 11 Class Diagram	23
Gambar 3. 12 Rancangan Halaman Login	24
Gambar 3. 13 Rancangan Dashboard.....	25
Gambar 3. 14 Rancangan Data Pengguna.....	25
Gambar 3. 15 Rancangan Data Review	26
Gambar 3. 16 Rancangan Hasil Analisa	26
Gambar 4. 1 Halaman Login.....	27
Gambar 4. 2 Halaman Dashboard	28
Gambar 4. 3 Halaman Admin Users	28
Gambar 4. 4 Halaman Review Users	29
Gambar 4. 5 Halaman Review	29
Gambar 4. 6 Halaman Algoritma Naïve Bayes.....	30
Gambar 4. 7 Halaman Algoritma KNN	30
Gambar 4. 8 Halaman Dashboard Pelanggan	31
Gambar 4. 9 Halaman Review List Pelanggan	31
Gambar 4. 10 Halaman Algoritma Naïve Bayes Pelanggan	32
Gambar 4. 11 Halaman Algoritma KNN Pelanggan.....	32
Gambar 4. 12 Diagram Alir KNN.....	33
Gambar 4. 13 Diagram Alir Naive Bayes	47