

**PENERAPAN SELEKSI FITUR *INFORMATION GAIN*
DALAM KLASIFIKASI ULASAN *TRAVELOKA* DI *GOOGLE*
PLAY STORE DENGAN *NAÏVE BAYES***



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2026

PROPOSAL TUGAS AKHIR

**PENERAPAN SELEKSI FITUR *INFORMATION GAIN* DALAM
KLASIFIKASI ULASAN *TRAVELOKA* DI *GOOGLE PLAY*
STORE DENGAN *NAÏVE BAYES***

Diajukan Untuk Memenuhi Persyaratan Guna Meraih Gelar Sarjana Komputer
Teknik Informatika Universitas Muhammadiyah Jember



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2026**

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR
PENERAPAN SELEKSI FITUR *INFORMATION GAIN* DALAM
KLASIFIKASI ULASAN *TRAVELOKA* DI *GOOGLE PLAY*
STORE* DENGAN *NAÏVE BAYES

Oleh:

Putri Noer Kumalasari

NIM. 2110651026

Telah Disetujui Bahwa Laporan Tugas Akhir Ini Untuk Diajukan Pada Sidang
Tugas Akhir Sebagai Salah Satu Syarat Kelulusan Dan Mendapat Gelar Sarjana
Di Universitas Muhammadiyah Jember

Disetujui Oleh:

Pembimbing I



Moh Dasuki, S.Kom., M.Kom
NIDN. 0722109103

Pembimbing II



Miftahur Rahman, S.Kom., M.Kom
NIDN. 0724039201

HALAMAN PENGESAHAN
PENERAPAN SELEKSI FITUR *INFORMATION GAIN* DALAM
KLASIFIKASI ULASAN *TRAVELOKA* DI *GOOGLE PLAY*
STORE* DENGAN *NAÏVE BAYES

Oleh:

Putri Noer Kumalasari

2110651026

Telah Mempertanggung Jawabkan Laporan Tugas Akhir pada Sidang Tugas
Akhir tanggal 5 Februari 2026 sebagai salah satu syarat kelulusan dan
Mendapatkan gelar Sarjana Komputer (S.Kom) Di Universitas Muhammadiyah
Jember

Disetujui Oleh:

Penguji I


Ginanjar Abdurrahman, S.Si., M.Pd
NIDN. 0714078704

Pembimbing I


Moh Dasuki, M.Kom
NIDN. 0722109103

Penguji II


Nur Qodariyah Fitriyah, S.T., M.Kom
NIDN. 0727097501

Pembimbing II


Miftahur Rahman, S.Kom., M.Kom
NIDN. 0724039201

**Mengesahkan,
Dekan Fakultas Teknik**


Prof. Dr. Ir. Muhtar, S.T., M.T., IPM
NIDN. 0010067301

**Mengetahui,
Ketua Program Studi Teknik
Informatika**


Rosita Yanuarti, S.Kom., M.Cs
NIDN. 0629018601

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : Putri Noer Kumalasari

NIM : 2110651026

Program Studi : Teknik Informatika

Judul Tugas Akhir : “Penerapan Seleksi Fitur *Information Gain* Dalam Klasifikasi
Ulasan *Traveloka* Di *Google Play Store* Dengan *Naïve Bayes*”

dengan ini menyatakan bahwa karya ilmiah yang saya susun adalah murni hasil pemikiran, penelitian, dan penyusunan saya sendiri. Segala sumber informasi yang berasal atau dikutip dari karya orang lain, baik berupa publikasi, artikel, buku, maupun sumber lainnya, telah disebutkan dengan jelas di dalam daftar pustaka sesuai dengan kaidah penulisan ilmiah.

Saya menyadari bahwa dalam proses penyusunan Tugas Akhir ini, saya memanfaatkan bantuan teknologi kecerdasan buatan (AI) seperti *ChatGPT/Grammarly/Alat AI lain* hanya sebatas untuk:

1. Membantu dalam penyusunan tata bahasa, parafrasa, atau perbaikan redaksi.
2. Memberikan inspirasi ide awal atau kerangka berpikir, yang selanjutnya saya kembangkan secara mandiri.
3. Membantu dalam pengecekan konsistensi format, ejaan, dan tata tulis.

Saya menegaskan bahwa tidak ada bagian dari karya ilmiah ini yang seluruhnya dibuat oleh AI tanpa keterlibatan pemikiran kritis saya sendiri. Tanggung jawab penuh atas isi, keaslian, dan kebenaran karya ilmiah ini ada pada diri saya.

Apabila di kemudian hari terbukti bahwa pernyataan ini tidak benar, maka saya bersedia menerima segala konsekuensi sesuai dengan peraturan yang berlaku di Universitas Muhammadiyah Jember.

Jember, 5 Februari 2026
Yang membuat pernyataan,



(Putri Noer Kumalasari)

MOTTO

“Kerja keras memang capek, tapi senyum setelah berhasil itu nggak ternilai.”

(Spongebob Squarepants)



KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur ke hadirat Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir berjudul “Penerapan Seleksi Fitur Information Gain dalam Klasifikasi Ulasan Traveloka di Google Play Store dengan Naïve Bayes.” Salawat dan salam semoga tercurah kepada Nabi Muhammad SAW.

Tugas Akhir ini terselesaikan berkat bantuan dan dukungan dari berbagai pihak, sehingga penulis mengucapkan terima kasih kepada:

1. Dr. Ir. Muhtar, S.T., M.T., IPM, selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
2. Rosita Yanuarti, S.Kom., M.Cs., selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
3. Moh. Dasuki, S.Kom., M.Kom. dan Miftahur Rahman, S.Kom., M.Kom., selaku dosen pembimbing atas arahan dan bimbingannya.
4. Nur Qodariyah Fitriyah, S.T., M.Kom. dan Ginanjar Abdurrahman, S.Si., M.Pd., selaku dosen penguji atas saran dan masukannya.
5. Kedua Orang tua saya dan keluarga yang senantiasa menjadi sumber doa, motivasi, dan dukungan.
6. Rekan-rekan seperjuangan atas motivasi dan kebersamaannya.
7. Semua pihak yang turut berperan, baik secara langsung maupun tidak langsung, dalam mendukung penyusunan Tugas Akhir ini.

Tugas Akhir ini masih memiliki keterbatasan, sehingga kritik dan saran yang membangun sangat diharapkan. Semoga karya ini bermanfaat bagi pembaca dan penelitian selanjutnya.

Terimakasih,
Wassalamu'alaikum warahmatullahi wabarakatuh.

DAFTAR ISI

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR.....	iii
HALAMAN PENGESAHAN.....	iv
LEMBAR PERNYATAAN KEASLIAN	v
MOTTO	vi
ABSTRAK	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Masalah	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Analisis Sentimen.....	6
2.2 <i>Traveloka</i>	6
2.3 <i>Bag Of Word</i>	7
2.4 <i>Information Gain</i>	8
2.4.1 Menghitung Entropi.....	9
2.4.2 Menghitung Entropi Kondisional.....	10
2.5 <i>Naïve Bayes Classifier</i>	10
2.6 <i>K-Fold Cross Validation</i>	11
2.7 <i>Confusion Matrix</i>	11
2.8 <i>Pyhton</i>	13
2.9 <i>Google Colab</i>	14
2.10 Penelitian Terdahulu.....	14
BAB III METODE PENELITIAN	17
3.1 Tahapan Penelitian	17
3.2 Pengumpulan Data	17
3.3 Pelabelan Data	17
3.4 <i>Preprocessing</i> Data	17
3.4.1 <i>Cleaning</i>	18
3.4.2 <i>Tokenization</i>	18

3.4.3 <i>Case Folding</i>	19
3.4.4 <i>Stopword Removal</i>	19
3.4.5 <i>Stemming</i>	20
3.5 Teknik <i>Bag of Word</i>	21
3.6 Seleksi Fitur <i>Information Gain</i>	22
3.7 Tanpa Seleksi Fitur <i>Information Gain</i>	27
3.8 Evaluasi Model	32
3.8.1 Evaluasi Model Dengan Seleksi Fitur <i>Information Gain</i>	32
3.8.2 Evaluasi Model Tanpa Seleksi Fitur <i>Information Gain</i>	33
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Pengumpulan Data	34
4.2 <i>Labeling Data</i>	37
4.2.1 <i>Labeling</i> Pakar Bahasa	37
4.2.2 <i>Labeling</i> Peneliti	38
4.3 <i>Preprocessing Data</i>	39
4.3.1 <i>Cleaning Data</i>	39
4.3.2 <i>Tokenizing</i>	40
4.3.3 <i>Case Folding</i>	40
4.3.4 <i>Stopword Removal</i>	41
4.3.5 <i>Stemming</i>	42
4.4 Seleksi Fitur <i>Information Gain</i>	42
4.5 Tanpa Seleksi Fitur <i>Information Gain</i>	46
4.6 Pemodelan <i>Naive Bayes</i> Untuk Seleksi Fitur <i>Information Gain</i>	47
4.7 Pemodelan <i>Naive Bayes</i> Tanpa Seleksi Fitur <i>Information Gain</i>	51
4.8 Evaluasi <i>Confusion Matrix</i>	54
4.8.1 Evaluasi <i>Confusion Matriks</i> Dengan Seleksi Fitur <i>Information Gain</i> ..	55
4.8.2 Evaluasi <i>Confusion Matriks</i> Tanpa Seleksi Fitur	58
BAB V KESIMPULAN DAN SARAN	63
5.1 Kesimpulan	63
5.2 Saran	64
DAFTAR PUSTAKA	65
Lampiran	67

DAFTAR GAMBAR

Gambar 2. 1 Contoh Data Ulasan	7
Gambar 2. 2 Memecah Kalimat Menjadi Per Kata	7
Gambar 3. 1 Tahapan Penelitian	17
Gambar 4. 1 Install <i>Google Play Scrapper</i>	34
Gambar 4. 2 Proses Seleksi Data	35
Gambar 4. 3 Proses Pemilihan Data	35
Gambar 4. 4 Proses Penyimpanan Data Hasil Scrapping	36
Gambar 4. 5 Proses Pengambilan Data	36
Gambar 4. 6 Data Hasil <i>Scrapping</i>	37
Gambar 4. 7 <i>Library</i> yang dibutuhkan	47
Gambar 4. 8 Data yang digunakan	48
Gambar 4. 9 Pemilihan fitur	48
Gambar 4. 10 Proses pembagian data	49
Gambar 4. 11 Proses pelatihan dan pengujian pada pemodelan	49
Gambar 4. 12 Evaluasi model	50
Gambar 4. 13 Penjabaran <i>confusion matriks</i>	50
Gambar 4. 14 <i>Import library</i> yang dibutuhkan	51
Gambar 4. 15 Data yang akan digunakan	51
Gambar 4. 16 Proses <i>Bag of word</i>	51
Gambar 4. 17 Pembagian data	52
Gambar 4. 18 Penentuan label untuk evaluasi <i>confusion matriks</i>	52
Gambar 4. 19 Pembagian data untuk pemodelan	53
Gambar 4. 20 Evaluasi pemodelan	53
Gambar 4. 21 Penjabaran <i>confusion matriks</i>	54
Gambar 4. 22 Menghitung rata-rata akurasi setiap fold	54
Gambar 4. 23 Perbandingan rata-rata akurasi	54
Gambar 4. 24 Fold 1 Seleksi Fitur	55
Gambar 4. 25 Fold 2 Seleksi Fitur	56
Gambar 4. 26 Fold 3 Seleksi Fitur	57
Gambar 4. 27 Fold 4 Seleksi Fitur	57
Gambar 4. 28 Fold 5 Seleksi Fitur	58
Gambar 4. 29 Fold 1 Tanpa Seleksi Fitur	59
Gambar 4. 30 Fold 2 Tanpa Seleksi Fitur	60
Gambar 4. 31 Fold 3 Tanpa Seleksi Fitur	60
Gambar 4. 32 Fold 4 Tanpa Seleksi Fitur	61
Gambar 4. 33 Fold 5 Tanpa Seleksi Fitur	62

DAFTAR TABEL

Tabel 2. 1 Jumlah Frekuensi Kata.....	8
Tabel 2. 2 <i>Confusion Matrix</i>	12
Tabel 2. 4 Penelitian Terdahulu	14
Tabel 3. 1 Proses <i>Cleaning</i>	18
Tabel 3. 2 Proses <i>Tokenization</i>	19
Tabel 3. 3 Proses <i>Case Folding</i>	19
Tabel 3. 4 Proses <i>Stopword Removal</i>	20
Tabel 3. 5 Proses <i>Stemming</i>	20
Tabel 3. 6 Data Ulasan Setelah Preprocessing.....	21
Tabel 3. 7 Memecah kalimat menjadi kata per kata	22
Tabel 3. 8 Frekuensi kata	22
Tabel 3. 9 Kategori Ulasan	23
Tabel 3. 10 Distribusi Kelas.....	23
Tabel 3. 11 Frekuensi Kata Terbanyak	Error! Bookmark not defined.
Tabel 3. 12 Keberadaan Fitur.....	24
Tabel 3. 13 Klasifikasi Fitur	25
Tabel 3. 14 Kategori Ulasan	27
Tabel 3. 15 Frekuensi Kata Ulasan Positif.....	28
Tabel 3. 16 Frekuensi Kata Ulasan Negatif	29
Tabel 3. 17 Klasifikasi <i>Confusion Matrix</i> Dengan <i>Information Gain</i>	32
Tabel 3. 18 Klasifikasi <i>Confusion Matrix</i> Tanpa <i>Information Gain</i>	33
Tabel 3. 19 Perbandingan Akurasi Evaluasi Model.....	33
Tabel 4. 1 Hasil Labeling Pakar Bahasa.....	37
Tabel 4. 2 <i>Library Labeling</i> Ulasan Otomatis	38
Tabel 4. 3 Hasil Labeling Otomatis Peneliti	39
Tabel 4. 4 Data Ulasan.....	39
Tabel 4. 5 Hasil <i>Cleaning</i> Data.....	39
Tabel 4. 6 <i>Library Tokenizing</i>	40
Tabel 4. 7 Hasil <i>Tokenizing</i>	40
Tabel 4. 8 Hasil <i>Case Folding</i>	41
Tabel 4. 9 <i>Library Stopword Removal</i>	41
Tabel 4. 10 Hasil <i>Stopword Removal</i>	41
Tabel 4. 11 <i>Library Stemming</i>	42
Tabel 4. 12 Hasil <i>Stemming</i>	42
Tabel 4. 13 Data Bahan Perhitungan	43
Tabel 4. 14 Skor <i>Information Gain</i> menggunakan 5 data ulasan.....	45
Tabel 4. 15 Hasil Perhitungan Keseluruhan <i>Information Gain</i>	46
Tabel 4. 16 Frekuensi Kata Tanpa Seleksi Fitur menggunakan 5 data ulasan.....	46
Tabel 4. 17 Hasil Perhitungan Keseluruhan Frekuensi Kata Tanpa Seleksi Fitur	47
Tabel 4. 19 Perolehan Akurasi Tiap Fold Dengan Seleksi Fitur	55
Tabel 4. 20 <i>Confusion Matrix</i> Fold 1	56
Tabel 4. 21 <i>Confusion Matrix</i> Fold 2	56
Tabel 4. 22 <i>Confusion Matrix</i> Fold 3	57
Tabel 4. 23 <i>Confusion Matrix</i> Fold 4	58
Tabel 4. 24 <i>Confusion Matrix</i> Fold 5	58
Tabel 4. 25 Perolehan Akurasi Tiap Fold Tanpa Seleksi Fitur.....	59

Tabel 4. 26 <i>Confusion Matrix</i> Fold 1	59
Tabel 4. 27 <i>Confusion Matrix</i> Fold 2	60
Tabel 4. 28 <i>Confusion Matrix</i> Fold 3	61
Tabel 4. 29 <i>Confusion Matrix</i> Fold 4	61
Tabel 4. 30 <i>Confusion Matrix</i> Fold 5	62

