

**PENGARUH PENGGUNAAN SELEKSI FITUR CHI SQUARE DAN  
INFORMATION GAIN TERHADAP PERFORMANSI ALGORITMA  
NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN  
APLIKASI *GOOGLE PLAY STORE*  
(STUDI KASUS : *APLIKASI MY F&B ID*)**

**SKRIPSI**



Oleh :

**NABILA AYU PUSPITA**

**NPM. 19081010006**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR**

**2024**

## LEMBAR PENGESAHAN SKRIPSI

**Judul** : PENGARUH PENGGUNAAN SELEKSI FITUR CHI SQUARE DAN INFORMATION GAIN TERHADAP PERFORMANSI ALGORITMA NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE PLAY STORE (STUDI KASUS : APLIKASI MY F&B ID)

**Oleh** : NABILA AYU PUSPITA

**NPM** : 19081010006

Telah Diseminarkan Dalam Ujian Skripsi Pada:

Hari Jumat, Tanggal 05 Januari 2024

Mengetahui

Dosen Pembimbing

1.

  
Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

2.

  
Agung Mustika Rizki, S.Kom, M.Kom

NIP. 19930725 202203 1 008

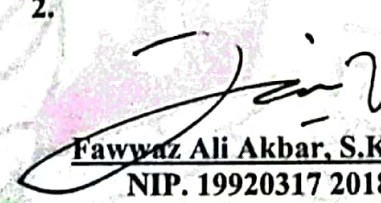
Dosen Penguji

1.

  
Dr. Rr. Ani Diah Rahajoe, ST., M.Cs

NIP. 19730512 200501 2003

2.

  
Fawwaz Ali Akbar, S.Kom, M.Kom

NIP. 19920317 201803 1002

Menyetujui

  
Dekan  
Fakultas Ilmu Komputer

  
Prof. Dr. Ir. Novirina Hendrasarie, M.T.

NIP. 19681126 199403 2 001

Koordinator Program Studi  
Informatika

  
Fetty Tri Anggraeny, S.Kom, M.Kom

NIP. 19820211 2021212 005

## SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa program studi Informatika UPN “Veteran” Jawa Timur, yang bertanda tangan dibawah ini:

Nama : Nabila Ayu Puspita

NPM : 19081010006

Menyatakan bahwa judul skripsi yang saya ajukan dan kerjakan berjudul:

**“PENGARUH PENGGUNAAN SELEKSI FITUR CHI SQUARE DAN INFORMATION GAIN TERHADAP PERFORMANSI ALGORITMA NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN APLIKASI *GOOGLE PLAY STORE* (STUDI KASUS : *APLIKASI MY F&B ID*)”**

Bukan merupakan plagiat dari skripsi/tugas akhir/penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lainnya.

Surabaya, 22 Januari 2024

Hormat Saya,



**Nabila Ayu Puspita**

**NPM. 19081010006**

## **KATA PENGANTAR**

Segenap puji syukur kehadiran Allah SWT dan shalawat senantiasa tercurahkan untuk Baginda Rasulullah Muhammad SAW yang menjadi teladan untuk umat muslim di dunia. Peneliti dapat menyelesaikan skripsi yang berjudul “PENGARUH PENGGUNAAN SELEKSI FITUR CHI SQUARE DAN INFORMATION GAIN TERHADAP PERFORMANSI ALGORITMA NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN APLIKASI GOOGLE PLAY STORE (STUDI KASUS : APLIKASI MY F&B ID)”.

Dukungan dan dorongan yang peneliti dapatkan selama pengerjaan penelitian serta dapat menyelesaikan laporan penelitian ini. Terima kasih peneliti ucapkan kepada para pihak yang telah turut secara langsung dan tidak langsung pada proses penyusunan laporan ini.

Peneliti menyadari masih ada kekurangan dalam penelitian dan penyusunan laporan skripsi berdasarkan keterbatasan kemampuan dan pengetahuan peneliti. Oleh karena ini, peneliti sangat menginginkan masukan, kritik dan saran yang membangun dan memperbaiki laporan ini menjadi lebih baik.

Surabaya, 22 Januari 2024

Penulis

## UCAPAN TERIMA KASIH

Puji dan syukur peneliti panjatkan kehadirat Allah SWT, atas berkah dan rahmat-Nya memberikan kekuatan, kesabaran, ilmu dan kesempatan untuk dapat menyelesaikan skripsi ini. Peneliti menyadari bahwa dalam penyusunan dan pengerjaan skripsi ini meliputi bantuan dan dukungan dari beberapa pihak. Maka pada kesempatan ini peneliti menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT, selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, S.T, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Ketua Program Studi Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku dosen pembimbing I yang telah meluangkan waktu dan tenaga dalam membimbing peneliti dengan baik dan memberikan masukan serta saran untuk dapat menyelesaikannya dengan tepat waktu.
5. Bapak Agung Mustika Rizki, S.Kom., M.Kom, selaku dosen pembimbing II yang telah meluangkan waktu dan tenaga dalam membimbing peneliti, memberi masukan dan saran dalam penyusunan skripsi menjadi baik dan benar.
6. Dosen Teknik Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang membantu dan memberi masukan dalam segala hal dalam pengerjaan skripsi ini.
7. Kedua orang tua saya Ayahanda tercinta Agus Suparman dan Ibunda tercinta Sulistini yang telah merawat, menjaga, mengorbankan waktu dan tenaga untuk anak-anaknya serta motivasi peneliti selama menyelesaikan skripsi ini.
8. Teman-teman Informatika angkatan 2019 yang telah membantu, mendukung dan berjuang dalam dunia perkuliahan.
9. Pihak-pihak yang tidak dapat peneliti ucapkan menjadi penyemangat peneliti dalam menyelesaikan skripsi ini.

Semoga Allah SWT memberikan balasan berupa kebaikan dunia dan akhirat kepada semua pihak yang telah peneliti sebutkan dalam menyelesaikan penelitian dan penyusunan laporan skripsi dengan baik.

Surabaya, 22 Januari 2024

Penulis

**PENGARUH PENGGUNAAN SELEKSI FITUR CHI SQUARE DAN INFORMATION GAIN TERHADAP PERFORMANSI ALGORITMA NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ULASAN APLIKASI *GOOGLE PLAY STORE* (STUDI KASUS : *APLIKASI MY F&B ID*)**

**Nama Mahasiswa : Nabila Ayu Puspita**

**NPM : 19081010006**

**Program Studi : Teknik Informatika**

**Dosen Pembimbing : Fetty Tri Anggraeny, S.Kom., M.Kom**

**Agung Mustika Rizki, S.Kom., M.Kom**

**Abstrak**

Di era teknologi yang semakin canggih mendorong semua aspek bidang kehidupan ikut dalam perkembangannya, salah satunya pada bidang maknaan atau kuliner. Kecanggihan hal tersebut ialah membuat perusahaan makanan berlomba-lomba untuk membuat aplikasi *mobile* untuk memberi kemudahan pada masyarakat terhadap pembelian produk yang mereka jual. Salah satu bentuk implementasi nyata dari hal tersebut dengan adanya Aplikasi My F&B ID. Pada *Google Play Store* selain dapat mengunduh aplikasi, masyarakat juga dapat memberi pendapat terhadap aplikasi tersebut. Pendapat yang berupa memberi ulasan masukkan, kritik, pujian, keluh kesah terhadap aplikasi. . Maka dalam penelitian ini akan dilakukan analisis sentimen terhadap ulasan aplikasi My F&B ID pada Google Playstore. Penelitian ini menggunakan algoritma Naïve Bayes Classifier dengan seleksi fitur Chi Square dan Information Gain dengan perbandingan dataset untuk data latih dan data uji sebesar 80:20, 70:30, 60:40. Perbandingan dataset dilakukan untuk mengetahui performansi nilai *accuracy* dari masing-masing model klasifikasi. Hasil dari penelitian diperoleh nilai *accuracy* tertinggi dihasilkan oleh model Klasifikasi Multinomial Naïve Bayes Classifier dengan Seleksi Fitur Chi Square ( $\alpha = 0.5$ ) dan Klasifikasi Multinomial Naïve Bayes Classifier tanpa Seleksi Fitur dengan nilai sebesar 85%.

**Kata kunci :** *naïve bayes classifier, analisis sentimen, ulasan, seleksi fitur, chi square, information gain*

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	ii
SURAT PERNYATAAN ANTI PLAGIAT .....	iii
KATA PENGANTAR .....	iv
UCAPAN TERIMA KASIH.....	v
Abstrak .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
DAFTAR KODE PROGRAM.....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Masalah .....	4
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 My F&B ID .....	8
2.3 Google Playstore .....	9
2.4 Analisis Sentimen.....	10
2.5 Preprocessing Text .....	10
2.6 TF-IDF.....	12
2.7 Chi Square .....	13



2.8	Information Gain .....	15
2.9	Naïve Bayes Classifier .....	17
2.10	Confusion Matrix .....	18
BAB III METODE PENELITIAN.....		21
3.1	Alur Penelitian.....	21
3.2	Studi Literatur.....	23
3.3	Pengumpulan Data .....	24
3.4	Preprocessing Text .....	24
3.5	Labelisasi Data .....	27
3.6	TF-IDF.....	28
3.7	Seleksi Fitur.....	30
3.7.1	Chi Square.....	31
3.7.2	Information Gain.....	33
3.8	Klasifikasi Model .....	36
3.9	Pengujian Sistem dan Evaluasi.....	39
BAB IV HASIL DAN PEMBAHASAN .....		40
4.1	<i>Scrapping</i> Dataset.....	40
4.2	Preprocessing Text Dataset .....	42
4.2.1	Case Folding.....	44
4.2.2	Cleaning .....	44
4.2.3	Tokenization.....	46
4.2.4	Normalisasi .....	47
4.2.5	Stopwords Removal .....	48
4.2.6	Stemming .....	50
4.3	Labelisasi Data .....	51

4.4	Pembobotan Kata .....	54
4.5	Seleksi Fitur.....	55
4.5.1	Seleksi Fitur Chi Square .....	55
4.5.2	Seleksi Fitur Information Gain.....	59
4.6	Pembagian Data Latih dan Data Uji.....	66
4.7	Model Klasifikasi Multinomial Naïve Bayes .....	67
4.8	Pengujian Sistem dan Evaluasi.....	70
4.8.1	Pengujian Multinomial Naïve Bayes Tanpa Seleksi Fitur .....	71
4.8.2	Pengujian Multinomial Naïve Bayes dengan Seleksi Fitur Chi Square 75	
4.8.3	Pengujian Multinomial Naïve Bayes dengan Seleksi Fitur Information Gain.....	83
BAB V KESIMPULAN DAN SARAN.....		94
5.1	Kesimpulan.....	94
5.2	Saran.....	95
DAFTAR PUSTAKA .....		96

## DAFTAR TABEL

Tabel 2. 1 Confusion Matrix .....	18
Tabel 3. 1 Contoh Dataset Case Folding.....	25
Tabel 3. 2 Contoh Tabel Tokenization.....	25
Tabel 3. 3 Contoh Dataset Normalisasi.....	26
Tabel 3. 4 Stopwords Removal .....	26
Tabel 3. 5 Contoh Dataset Stemming .....	27
Tabel 3. 6 Contoh Labelisasi Data menggunakan InSet .....	28
Tabel 3. 7 Contoh Dataset Penelitian.....	29
Tabel 3. 8 TF-IDF pada tiap term di dokumen .....	30
Tabel 3. 9 Contoh Dataset Penelitian.....	31
Tabel 3. 10 Kemunculan dan Ketidakmunculan Term pada Tiap Kelas .....	31
Tabel 3. 11 Hasil Perhitungan Term Pada Tiap Kelas .....	33
Tabel 3. 12 Contoh Dataset Penelitian.....	33
Tabel 3. 13 Kemunculan Fitur Pada Tiap Kelas .....	34
Tabel 3. 14 Fitur dan Nilai Information Gain .....	36
Tabel 3. 15 Contoh Data Latih dan Data Uji Penelitian .....	36
Tabel 3. 16 Probabilitas Tiap Kata Terhadap Kelas Sentimen .....	37
Tabel 4. 1 Data Latih Dan Data Uji .....	71
Tabel 4. 2 Perbandingan Performansi tiap Nilai taraf nyata ( $\alpha$ ) .....	75
Tabel 4. 3 Perbandingan Accuracy Naive Bayes Classifier dengan Fitur Terpilih dan Fitur Terbuang Hasil Proses Chi Square .....	83
Tabel 4. 4 Perbandingan Performansi tiap Nilai Threshold.....	83
Tabel 4. 5 Perbandingan Accuracy Naive Bayes Classifier dengan Fitur Terpilih dan Fitur Terbuang Hasil Proses Information Gain .....	91
Tabel 4. 6 Perbanding Nilai Accuracy Pada Klasifikasi .....	92

## DAFTAR GAMBAR

Gambar 2. 1 Logo My F&B ID.....	8
Gambar 2. 2 Tampilan Aplikasi My F&b ID pada Android .....	9
Gambar 3. 1 Alur Metode Penelitian .....	21
Gambar 3. 2 Alur Preprocessing .....	24
Gambar 3. 3 Perhitungan TF-IDF .....	28
Gambar 4. 1 Mengunduh Scrapped Data.....	42
Gambar 4. 2 Mengunggah File ke Google Colab .....	43
Gambar 4. 3 File dataset yang telah diunggah .....	43
Gambar 4. 4 Dataset Hasil Case Folding .....	44
Gambar 4. 5 Dataset Hasil Cleaning .....	46
Gambar 4. 6 Dataset Hasil Tokenization .....	47
Gambar 4. 7 Dataset Hasil Normalisasi .....	48
Gambar 4. 8 Dataset Hasil Stopwords Removal.....	49
Gambar 4. 9 Dataset Hasil Stemming .....	51
Gambar 4. 10 Dataset Hasil Labelisasi Data.....	54
Gambar 4. 11 Fitur dari Hasil Seleksi Fitur Chi Sqaure .....	58
Gambar 4. 12 Fitur Terbuang Hasil Seleksi Fitur Chi Square .....	58
Gambar 4. 13 Fitur Terpilih Hasil Proses Information Gain .....	62
Gambar 4. 14 Fitur Terbuang Hasil Proses Information Gain .....	65
Gambar 4. 15 Classification Report Skema Pengujian 1 .....	72
Gambar 4. 16 Confusion Matrix Skema Pengujian 1 .....	72
Gambar 4. 17 Classification Report Skema Pengujian 2 .....	73
Gambar 4. 18 Confusion Matrix Skema Pengujian 2 .....	73
Gambar 4. 19 Classification Report Skema Pengujian 3 .....	74
Gambar 4. 20 Confusion Matrix Skema Pengujian 3 .....	74
Gambar 4. 21 Classification Report Skema Pengujian 4 .....	76
Gambar 4. 22 Confusion Matrix Skema Pengujian 4 .....	76
Gambar 4. 23 Classification Report Skema Pengujian 5 .....	77
Gambar 4. 24 Confusion Matrix Skema Pengujian 5 .....	77
Gambar 4. 25 Classification Report Skema Pengujian 6 .....	78

Gambar 4. 26 Confusion Matrix Skema Pengujian 6 .....	79
Gambar 4. 27 Classification Report Skema Pengujian 7 .....	79
Gambar 4. 28 Confusion Matrix Skema Pengujian 7 .....	80
Gambar 4. 29 Classification Report Skema Pengujian 8 .....	81
Gambar 4. 30 Confusion Matrix Skema Pengujian 8 .....	81
Gambar 4. 31 Classification Report Skema Pengujian 9 .....	82
Gambar 4. 32 Confusion Matrix Skema Pengujian 9 .....	82
Gambar 4. 33 Classification Report Skema Pengujian 10 .....	84
Gambar 4. 34 Confusion Matrix Skema Pengujian 10 .....	85
Gambar 4. 35 Classification Report Skema Pengujian 11 .....	85
Gambar 4. 36 Confusion Matrix Skema Pengujian 11 .....	86
Gambar 4. 37 Classification Report Skema Pengujian 12 .....	87
Gambar 4. 38 Confusion Matrix Skema Pengujian 12 .....	87
Gambar 4. 39 Classification Report Skema Pengujian 13 .....	88
Gambar 4. 40 Confusion Matrix Skema Pengujian 13 .....	88
Gambar 4. 41 Classification Report Skema Pengujian 14 .....	89
Gambar 4. 42 Confusion Matrix Skema Pengujian 14 .....	89
Gambar 4. 43 Classification Report Skema Pengujian 15 .....	90
Gambar 4. 44 Confusion Matrix Skema Pengujian 15 .....	91

## DAFTAR KODE PROGRAM

Kode Program 4. 1 Scrapping Dataset F&B ID .....	41
Kode Program 4. 2 Mengimpor Dataset .....	43
Kode Program 4. 3 Membaca Dataset.....	43
Kode Program 4. 4 Case Folding .....	44
Kode Program 4. 5 Cleaning .....	45
Kode Program 4. 6 Normalisasi .....	48
Kode Program 4. 7 Stopwords Removal.....	49
Kode Program 4. 8 Stemming.....	51
Kode Program 4. 9 Kode Program Labelisasi Data .....	54
Kode Program 4. 10 Pembobotan Kata.....	55
Kode Program 4. 11 Proses Chi Square dan Tabel Fitur dengan nilai p-value $\leq$ nilai taraf nyata ( $\alpha$ ) .....	57
Kode Program 4. 12 Class Information Gain.....	60
Kode Program 4. 13 Seleksi Fitur Information Gain Terpilih .....	62
Kode Program 4. 14 Fitur Terbuang dari Proses Information Gain.....	64
Kode Program 4. 16 Memilih Fitur Yang Digunakan.....	65
Kode Program 4. 17 Label Encoding .....	66
Kode Program 4. 18 Pembagian Dataset.....	67
Kode Program 4. 19 Multinoial Naïve Bayes .....	69
Kode Program 4. 20 Melatih Model Klasifikasi .....	70
Kode Program 4. 21 Pengujian Model Klasifikasi terhadap $X_{test}$ .....	70
Kode Program 4. 22 Classification Report dan Confusion Matrix .....	71