



**ANALISIS SENTIMEN TERHADAP LAYANAN
TRANSJAKARTA PADA MEDIA SOSIAL *INSTAGRAM*
MENGUNAKAN METODE *NAÏVE BAYES* DAN SELEKSI
FITUR *INFORMATION GAIN***

SKRIPSI

Ivtytah Ein

1810511010

UNIVERSITAS PEMBANGUNAN VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI SARJANA INFORMATIKA

2022



**ANALISIS SENTIMEN TERHADAP LAYANAN
TRANSJAKARTA PADA MEDIA SOSIAL *INSTAGRAM*
MENGUNAKAN METODE *NAÏVE BAYES* DAN SELEKSI
FITUR *INFORMATION GAIN***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

Ivtytah Ein

1810511010

UNIVERSITAS PEMBANGUNAN VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI SARJANA INFORMATIKA

2022

PERNYATAAN ORISINALITAS

Tugas Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun dirujuk saya nyatakan dengan benar.

Nama : Ivtytah Ein

NIM : 1810511010

Tanggal : 19 Juli 2022

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 19 Juli 2022

Yang Menyatakan,



(Ivtytah Ein)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Ivtytah Ein
NIM : 1810511010
Fakultas : Ilmu Komputer
Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Analisis Sentimen Terhadap Layanan Transjakarta Pada Media Sosial Instagram Menggunakan Metode *Naïve Bayes* dan Seleksi Fitur *Information Gain*.

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti di Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 19 Juli 2022

Yang Menyatakan,



(Ivtytah Ein)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Ivtytah Ein

NIM : 1810511010

Program Studi : S1 Informatika 2018

Judul : Analisis Sentimen Terhadap Layanan Transjakarta Pada Media Sosial *Instagram* Menggunakan Metode *Naïve Bayes* dan Seleksi Fitur *Information Gain*.

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Jayanta, S.Kom., M.Si.
Penguji 1

Nurul Chamidah, S.Kom., M.Kom.
Penguji 2

Iin Ernawati, S.Kom., M.Si.
Pembimbing 1

Yuni Widiastiwi, S.Kom., M.Si.
Pembimbing 2



Dr. Ermatita, M.Kom.
Dekan

Desta Sandya Prasvita, S.Kom., M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal pengesahan : 19 Juli 2022



**ANALISIS SENTIMEN TERHADAP LAYANAN TRANSJAKARTA PADA
MEDIA SOSIAL *INSTAGRAM* MENGGUNAKAN METODE *NAÏVE*
BAYES DAN SELEKSI FITUR *INFORMATION GAIN***

IVTYTAH EIN

ABSTRAK

PT. Transportasi Jakarta (Transjakarta) menanggapi kebijakan Pemerintah Provinsi DKI Jakarta mengenai pembatasan pada moda transportasi dengan membatasi armada bus, rute perjalanan, jam operasional dan kapasitas angkut. Pada tanggal 22 Oktober 2021, Transjakarta mengumumkan layanan akan kembali normal. Walaupun layanan kembali beroperasi dengan normal masih banyak pengguna transjakarta yang menuangkan kritik dan opini yang berkaitan dengan layanan transjakarta pada akun *instagram* transjakarta. Penelitian ini bertujuan untuk membangun model klasifikasi sentimen dengan menggunakan metode *naïve bayes* dan seleksi fitur *information gain* terhadap opini masyarakat terkait pelayanan transjakarta di media sosial *instagram*. Data komentar dibagi menjadi kelas positif dan kelas negatif berdasarkan pelabelan manual yang dilakukan oleh 3 *annotator* yang nantinya data tersebut akan dilakukan *pre-processing*, pembobotan TF-IDF, seleksi fitur, serta pembagian data sebesar 70% data latih dengan penerapan metode SMOTE dan 30% data uji sebelum masuk ke tahap pemodelan. Pada penelitian ini terdapat dua model yaitu model klasifikasi *naïve bayes* tanpa seleksi fitur dan dengan seleksi fitur. Hasil evaluasi untuk model klasifikasi *naïve bayes* memperoleh akurasi sebesar 81,42%, *recall* sebesar 69,64%, *precision* sebesar 63,93%, dan *specificity* sebesar 85,71%. Sedangkan hasil model klasifikasi *naïve bayes* dengan *information gain* memperoleh akurasi sebesar 86,66%, *recall* sebesar 71,42%, *precision* sebesar 76,92%, dan *specificity* sebesar 92,20%.

Kata kunci : Sentimen, *Instagram*, Transjakarta, *Naïve Bayes*, *Information Gain*.

**SENTIMENT ANALYSIS OF TRANSJAKARTA ON SOCIAL MEDIA
INSTAGRAM USING NAÏVE BAYES METHOD AND INFORMATION
GAIN FEATURE SELECTION**

IVTYTAH EIN

ABSTRACT

PT. Jakarta Transportation (Transjakarta) responds to the DKI Jakarta Provincial Government's policy regarding restrictions on transportation modes by limiting bus fleets, travel routes, operating hours and transport capacity. On October 22, 2021, Transjakarta announced that services would return to normal. Even though the service is operating normally again, there are still many transjakarta users who express criticism and opinions related to transjakarta services on their instagram account. This study aims to build a sentiment classification model using the naïve bayes method and selection of information gain features on public opinion regarding transjakarta services on Instagram social media. Comment data is divided into positive class and negative class based on manual labeling carried out by 3 annotators which later on the data will be pre-processed, TF-IDF weighting, feature selection, and data sharing of 70% training data using the SMOTE method and 30% test data before entering the modeling stage. In this study, there are two models, namely the nave Bayes classification model without feature selection and with feature selection. The evaluation results for the nave Bayes classification model obtained an accuracy of 81.42%, recall of 69.64%, precision of 63.93%, and specificity of 85.71%. While the results of the nave Bayes classification model with information gain obtained an accuracy of 86.66%, recall of 71.42%, precision of 76.92%, and specificity of 92.20%..

Keywords: *Sentiment, Instagram, Naïve Bayes, Feature Selection, Information Gain.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan atas berkat, rahmat dan karunia-Nya kepada Allah SWT, penulis dapat menyelesaikan Tugas Akhir/skripsi yang berjudul “**Analisis Sentimen Terhadap Layanan Transjakarta Pada Media Sosial Instagram Menggunakan Metode Naïve Bayes dan Seleksi Fitur Information Gain.**” dengan baik.

Dalam penyelesaian tugas akhir ini, penulis tidak lupa mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan masukan kepada penulis, pihak-pihak tersebut adalah:

1. Kedua orang tua penulis, (ayah), (ibu), serta keluarga yang selalu memberi dorongan dan doa sehingga dapat menyelesaikan proposal ini.
2. Ibu Iin Ernawati S.Kom., M.Si dan Yuni Widiastiwi, S.Kom., Msi., selaku dosen pembimbing yang telah memberikan saran dan masukan yang sangat bermanfaat.
3. Ibu Dr. Ermatita, M. Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Bapak Desta Sandya Prasvita, S.Kom., M.Kom.selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta.
5. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta terima kasih atas ilmu-ilmu yang selama ini sudah diajarkan.
6. Erisa, Aliffia, Luthfiyah, Tya, Albet, Ismail, Bobby, Hilda, Haiqal, Hasan, Ester, Mika, Feby selaku teman yang mendukung dan memberikan semangat.
7. Teman-teman Fakultas Ilmu Komputer angkatan 2018 dan seluruh pihak terkait yang membantu menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu, penulis ucapkan terima kasih.

Akhir kata, semoga tugas akhir/ skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 19 Juli 2021

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR SIMBOL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Luaran Yang Diharapkan	4
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Data Mining</i>	6
2.1.1 Definisi <i>Data Mining</i>	6
2.1.2 Tahapan <i>Data Mining</i>	7
2.1.3 Pengelompokan <i>Data Mining</i>	8
2.2 <i>Text Mining</i>	9
2.2.1 <i>Preprocessing</i>	10
2.3 <i>Fleiss Kappa</i>	11
2.4 Pembobotan TF-IDF	13
2.5 <i>Information Gain</i>	13
2.6 Analisis Sentimen	14

2.7	<i>Instagram</i>	15
2.8	Transjakarta.....	15
2.9	Metode <i>Naïve Bayes</i>	15
2.10	<i>Confusion Matrix</i>	18
2.10.1	Akurasi.....	18
2.10.2	<i>Recall</i>	19
2.10.3	<i>Precision</i>	19
2.10.4	<i>Specificity</i>	19
2.11	Penelitian Relevan.....	19
BAB III METODOLOGI PENELITIAN		23
3.1	Kerangka Pikir	23
3.1.1	Identifikasi Masalah	23
3.1.2	Studi Literatur	23
3.1.3	Pengumpulan Data	24
3.1.4	Pelabelan Data	25
3.1.5	<i>Preprocessing</i>	25
3.1.6	Pembobotan TF-IDF	27
3.1.7	Seleksi Fitur	27
3.1.8	Pembagian Data	28
3.1.9	Klasifikasi	29
3.1.10	Evaluasi	29
3.2	Alat Bantu Penelitian	30
3.3	Jadwal Penelitian	31
BAB IV HASIL DAN PEMBAHASAN		32
4.1	Pengumpulan Data	32
4.2	Pelabelan Data	33
4.3	<i>Pre-processing</i>	38
4.3.1	<i>Case Folding</i>	39
4.3.2	<i>Data Cleaning</i>	39
4.3.3	<i>Normalization</i>	40
4.3.4	<i>Stopword Removal</i>	41
4.3.5	<i>Stemming</i>	42
4.3.6	<i>Tokenizing</i>	43
4.4	Pembobotan TF-IDF	44

4.5	Seleksi Fitur	46
4.5.1	Pemisahan Fitur Sesuai Dengan Label	47
4.5.2	Perhitungan Nilai Entropi <i>Record</i> dan Total	48
4.5.3	Perhitungan Nilai Entropi Fitur	49
4.5.4	Perhitungan Nilai <i>Information Gain</i>	49
4.5.5	Pengurutan Fitur	51
4.5.6	Pengambilan Fitur	52
4.6	Klasifikasi	53
4.6.1	Pembagian Data	53
4.6.2	Pemodelan <i>Naïve Bayes</i>	54
4.6.3	Pemodelan <i>Naïve Bayes</i> dan <i>Information Gain</i>	61
4.7	Evaluasi	68
4.7.1	Evaluasi Model <i>Naïve Bayes</i>	68
4.7.2	Evaluasi Model <i>Naïve Bayes</i> dan <i>Information Gain</i>	69
4.7.3	Evaluasi Perbandingan Model Klasifikasi	71
4.8	Visualisasi	71
4.8.1	Visualisasi <i>Wordcloud</i> Sentimen	71
4.8.2	Visualisasi <i>Wordcloud</i> Sentimen Positif	72
4.8.3	Visualisasi <i>Wordcloud</i> Sentimen Negatif	72
4.8.4	Perbandingan Sentimen Positif dan Negatif	73
BAB V PENUTUP		74
5.1	Kesimpulan	74
5.2	Saran	75
DAFTAR PUSTAKA		76
RIWAYAT HIDUP		80
LAMPIRAN		81
Lampiran 1. Kamus Normalisasi		82
Lampiran 2. Kamus <i>Stopword</i>		95
Lampiran 3. Persentase Turnitin		102




DAFTAR GAMBAR

Gambar 2. 1 Data Mining Sebagai Proses Dalam KDD (Han et al., 2012).....	7
Gambar 3. 1 Kerangka Pikir	23
Gambar 4. 1 Data Sebelum Pandemi Layanan Transjakarta Normal	32
Gambar 4. 2 Data Saat Pandemi Layanan Trasnjakarta Normal	33
Gambar 4. 3 Wordcloud Sentimen	71
Gambar 4. 4 Wordcloud Sentimen Positif.....	72
Gambar 4. 5 Wordcloud Sentimen Negatif	72
Gambar 4. 6 Visualisasi Hasil Pelabelan Data Sentimen	73

DAFTAR TABEL

Tabel 2. 1 Tabel Skala Kappa Value	12
Tabel 2. 2 Tabel Confusion Matrix	18
Tabel 4. 1 Penghapusan Data Tidak Relevan	33
Tabel 4. 2 Hasil Pelabelan Data Sebelum Pandemi	34
Tabel 4. 3 Hasil Pelabelan Data Saat Pandemi	34
Tabel 4. 4 Sampel Data Untuk Kappa Value	35
Tabel 4. 5 Case Folding.....	39
Tabel 4. 6 Data Cleaning.....	40
Tabel 4. 7 Normalization.....	41
Tabel 4. 8 Stopword Removal.....	42
Tabel 4. 9 Stemming	43
Tabel 4. 10 Tokenizing.....	43
Tabel 4. 11 Sampel Data Pembobotan TF-IDF	44
Tabel 4. 12 Hasil Pembobotan TF-IDF	45
Tabel 4. 13 Sampel Data Seleksi Fitur	46
Tabel 4. 14 Hasil Perhitungan Information Gain Pada Sampel Data	49
Tabel 4. 15 Hasil Pengurutan Fitur	51
Tabel 4. 16 Hasil Pengambilan Fitur.....	52
Tabel 4. 17 Pembagian Data Naive Bayes	53
Tabel 4. 18 Pembagian Data Naive Bayes dan Information Gain.....	53
Tabel 4. 19 Penerapan Metode SMOTE	54
Tabel 4. 20 Sampel Data Model Naive Bayes.....	54
Tabel 4. 21 Nilai TF-IDF Sampel Data Model Naive Bayes	55
Tabel 4. 22 Nilai Probabilitas Data Latih Model Naive Bayes	58
Tabel 4. 23 Sampel Data Uji Model Naive Bayes.....	59
Tabel 4. 24 Hasil TF-IDF Sampel Data Uji Model Naive Bayes.....	59
Tabel 4. 25 Sampel Data Latih Model NB dan IG	61
Tabel 4. 26 Nilai TF-IDF Data Latih Model NB dan IG.....	62
Tabel 4. 27 Nilai Probabilitas Data Latih Model NB dan IG	64
Tabel 4. 28 Sampel Data Uji Model NB dan IG	65
Tabel 4. 29 Hasil TF-IDF Sampel Data Uji Model NB dan IG	65
Tabel 4. 30 Hasil Perhitungan Model Klasifikasi	68
Tabel 4. 31 Evaluasi Model Naive Bayes	69
Tabel 4. 32 Confusion Matrix Model NB dan IG	70
Tabel 4. 33 Hasil Evaluasi Perbandingan Model Klasifikasi	71

DAFTAR SIMBOL

No.	Simbol	Nama Simbol	Keterangan
1.		Simbol Terminator	Simbol yang digunakan untuk mulai dan akhir dari suatu kegiatan.
2.		Simbol Proses	Simbol yang digunakan untuk menggambarkan suatu proses kegiatan.
3.		Simbol arah data atau arus data	Simbol yang digunakan sebagai penunjuk arah aliran serta arus data pada suatu proses kegiatan.