

**PENERAPAN PRINCIPAL COMPONENT ANALYSIS PADA
ANALISIS SENTIMEN MENGGUNAKAN MULTINOMIAL
NAÏVE BAYES**

(Studi Kasus : Pelayanan Publik Kereta Api Lokal DAOP 8)

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data**



Disusun Oleh:

RISNALDY NOVENDRA IRAWAN

20083010017

**PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
SURABAYA
2024**

LEMBAR PENGESAHAN

PENERAPAN PRINCIPAL COMPONENT ANALYSIS PADA ANALISIS SENTIMEN MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

(Studi Kasus: Pelayanan Publik Kereta Api Lokal DAOP 8)

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data pada : Selasa, 14 Mei 2024

Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya

Oleh:

RISNALDY NOVENDRA IRAWAN

NPM. 20083010017

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Penguji 2

Wahyu Syaifulah J.S., S.Kom., M.Kom.

Aviolla Terza Damaliana, S.Si., M.Stat.

NIP. 198608252021211003

NIP. 199408022022032015

Pembimbing 1

Pembimbing 2

Kartika Maulida Hindrayani, S.Kom., M.Kom.

Dr. Ir. Mohammad Idhom, SP., S.Kom., MT.

NIP. 199209092022032009

NIP. 198303102021211006

Fakultas Ilmu Komputer
Dekan,

Mengetahui,

Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Prof. Dr. Ir. Novirina Hendrasarie, MT.

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.

NIP. 196811261994032001

NIP. 198012052005011002

Surabaya, Mei, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Risnaldy Novendra Irawan
NPM : 20083010017
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

PENERAPAN PRINCIPAL COMPONENT ANALYSIS PADA ANALISIS SENTIMEN MENGGUNAKAN MULTINOMIAL NAÏVE BAYES (Studi Kasus: Pelayanan Publik Kereta Api Lokal DAOP 8)

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 28 Mei 2024

Hormat Saya



Risnaldy Novendra Irawan
NPM. 20083010017

KATA PENGANTAR

Puji dan syukur kehadiran ALLAH SWT, atas limpahan Rahmat serta Kasih Sayang-Nya sehingga penulis dapat menyelesaikan Proposal Skripsi yang merupakan persyaratan dalam menyelesaikan mata kuliah Seminar Proposal pada Program Studi S1 Sains Data di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak, dan dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Allah SWT yang telah memberikan kekuatan rahmat, serta hidayah kepada penulis, hingga menyelesaikan skripsi ini sampai selesai
2. Mama Susilawati dan Adek Celia yang telah memberikan support materi dan doa agar skripsi ini bisa terselesaikan
3. Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom selaku dosen pembimbing 1 yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi
4. Bapak Dr. Ir. Mohammad Idhom, SP., S.Kom., MT selaku dosen pembimbing 2 yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi
5. Seluruh dosen Program Studi Sains Data atas ilmu-ilmu yang di berikan selama masa perkuliahan.
6. Teman-teman dekat yang selalu memberikan support dan arahan dalam mengerjakan tugas sampai skripsi ini terselesaikan

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam Proposal Skripsi ini, namun penulis berharap semoga Proposal Skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu sains data.

Surabaya, 28 Mei 2024

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Dasar Teori.....	6
2.2. Penelitian Terdahulu	15
BAB III METODOLOGI.....	18
3.1. Variabel Penelitian dan Sumber Data	18
3.2. Langkah Analisis.....	19
3.3. Diagram Alir Penelitian	20
3.4. Jadwal Penelitian.....	30
BAB IV HASIL DAN PEMBAHASAN	31

4.1.	Persiapan Data.....	31
4.1.1.	Pengambilan Data	31
4.1.2.	Pelabelan Data.....	32
4.2.	<i>Preprocessing Data</i>	33
4.2.1.	<i>Cleaning</i>	34
4.2.2.	<i>Case Folding</i>	34
4.2.3.	<i>Tokenizing</i>	35
4.2.4.	<i>Stopword Removal</i>	36
4.2.5.	<i>Stemming</i>	37
4.3.	<i>Label Encoding</i>	38
4.4.	Distribusi Data	39
4.5.	<i>Wordcloud</i>	40
4.6.	Pembobotan Data	42
4.7.	Seleksi Fitur	43
4.7.1.	Seleksi Fitur Label Manual	44
4.7.2.	Seleksi Fitur Label Prediksi	46
4.8.	Model <i>Multinomial Naïve Bayes</i>	47
4.8.1.	Pengujian Menggunakan <i>Multinomial Naïve Bayes</i>	47
4.9.	Evaluasi Model.....	48
4.9.1.	<i>Confusion Matrix</i>	49
4.9.2.	Skenario Eksperimen	49
4.9.3.	Analisis Eksperimen Pengujian.....	58
BAB V PENUTUP.....		66
5.1.	Kesimpulan	66
5.2.	Saran.....	68
DAFTAR PUSTAKA		69
LAMPIRAN.....		72
BIODATA PENULIS		95

DAFTAR GAMBAR

Gambar 3. 1. Diagram Alir Penelitian	20
Gambar 3. 2. Diagram Alir <i>Preprocessing</i> Data.....	22
Gambar 3. 3. Diagram Alir <i>TF – IDF</i>	25
Gambar 4. 1. Dataset Mentah.....	31
Gambar 4. 2. Hasil Analisis Sentimen menggunakan <i>library</i> <i>SentimentIntensityAnalyzer</i>	33
Gambar 4. 3. Hasil Menghapus Kolom Ulasan Bahasa Inggris dan <i>Compound</i> ..	33
Gambar 4. 4. Hasil <i>Label Encoding</i>	39
Gambar 4. 5. Hasil Distribusi Data	40
Gambar 4. 6. Grafik Distribusi Data	40
Gambar 4. 7. <i>WordCloud</i> Positif.....	41
Gambar 4. 8. <i>WordCloud</i> Negatif	42
Gambar 4. 9. Hasil Pembobotan <i>TF-IDF</i>	43
Gambar 4. 10. Grafik Hasil Penentuan PC terbaik dengan <i>Labeling</i> Secara Manual	45
Gambar 4. 11. Hasil Seleksi Fitur Terhadap Data <i>Labeling</i> Manual.....	45
Gambar 4. 12. Grafik Hasil Penentuan PC terbaik dengan <i>Labeling</i> Prediksi	46
Gambar 4. 13. Hasil Seleksi Fitur Terhadap Data Label Prediksi	47
Gambar 4. 14. <i>Classification Report</i> Pengujian <i>Multinomial Naïve Bayes</i> dengan Label Manual	50
Gambar 4. 15. <i>Classification Report</i> Pengujian <i>Multinomial Naïve Bayes</i> dengan Label Prediksi.....	51
Gambar 4. 16. <i>Classification Report</i> Hasil Pengujian <i>Multinomial Naive Bayes</i> PCA 827 Label Manual.....	54
Gambar 4. 17. <i>Classification Report</i> Hasil Pengujian <i>Multinomial Naive Bayes</i> PCA 114 Label Prediksi.....	55
Gambar 4. 18. <i>Classification Report</i> Hasil Pengujian <i>Multinomial Naive Bayes</i> PCA 472 Label Manual.....	56
Gambar 4. 19. <i>Classification Report</i> Hasil Pengujian <i>Multinomial Naive Bayes</i> PCA 1185 Label Prediksi.....	58

DAFTAR TABEL

Tabel 2. 1 Tabel <i>Confusion Matrix</i>	14
Tabel 2. 2 . Penelitian terdahulu.....	16
Tabel 3. 1. Dataset Ulasan Pelanggan yang Menggunakan Moda transportasi Kereta Api Subsidi	18
Tabel 3. 2. Tabel Hasil TF - IDF.....	26
Tabel 3. 3. Tabel Klasifikasi Hasil Uji.....	29
Tabel 3. 4. Jadwal Penelitian.....	30
Tabel 4. 1. Tabel Hasil <i>Cleaning Data</i>	34
Tabel 4. 2. Tabel Hasil <i>Case Folding</i>	35
Tabel 4. 3. Tabel Hasil Tokenisasi.....	36
Tabel 4. 4. Tabel Hasil <i>Stopword Removal</i>	37
Tabel 4. 5. Tabel Hasil <i>Stemming</i>	38
Tabel 4. 6. Tabel Eksperimen Klasifikasi <i>Multinomial Naive Bayes</i> Tanpa PCA	49
Tabel 4. 7. Hasil Pengujian <i>Multinomial Naive Bayes</i> Label Manual Tanpa PCA	50
Tabel 4. 8. Hasil Pengujian <i>Multinomial Naive Bayes</i> Label Prediksi Tanpa PCA	51
Tabel 4. 9. Skenario Eksperimen <i>Multinomial Naive Bayes</i> Menggunakan PCA	53
Tabel 4. 10. Hasil Pengujian PCA 827 <i>Multinomial Naive Bayes</i> Label Manual	53
Tabel 4. 11. Hasil Pengujian PCA 114 <i>Multinomial Naive Bayes</i> Label Prediksi	55
Tabel 4. 12. Tabel Hasil Eksperimen Pengujian	59

DAFTAR LAMPIRAN

Lampiran 1. Hasil Uji Plagiasi	72
Lampiran 2. Data Penelitian.....	83
Lampiran 3. <i>Source Code</i>	83
Lampiran 4. Surat Pengantar Penelitian Skripsi Permohonan Penelitian	94

ABSTRAK

PENERAPAN PRINCIPAL COMPONENT ANALYSIS PADA ANALISIS SENTIMEN MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

(Studi Kasus: Pelayanan Publik Kereta Api Lokal DAOP 8)

Nama Mahasiswa / NPM : Risnaldy Novendra Irawan / 20083010017
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim
Dosen Pembimbing 1 : Kartika Maulida Hindrayani, S.Kom., M.Kom
Dosen Pembimbing 2 : Dr. Ir. Mohammad Idhom, SP., S.Kom., MT.

Abstrak

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap pengguna layanan kereta api subsidi wilayah daerah operasional 8 Surabaya. Metode yang digunakan untuk analisis yaitu *Naïve Bayes Classifier* dengan tujuan mengetahui pengaruh seleksi fitur *Principal Component Analysis* terhadap algoritma *Multinomial Naïve Bayes*. Tahap *preprocessing* data yang dilakukan termasuk pembersihan, pembobotan, dan pembagian data sehingga didapatkan hasil sebanyak 1123 data dengan dua kelas yaitu kelas positif dan negatif yang kemudian dilakukan *processing* data berupa pencaian fitur terbaik PCA dan melakukan klasifikasi. Hasil pengolahan data dengan *Multinomial Naïve Bayes* menggunakan pelabelan manual tanpa seleksi fitur *Principal Component Analysis* memiliki kinerja yang lebih akurat dibandingkan dengan menggunakan seleksi fitur dengan akurasi yang didapatkan selama proses pengujian mencapai 80%. Skenario diatas juga menghasilkan data benar terprediksi sebesar 139 untuk positif dan 40 untuk negatif. Hasil lainnya terhadap data menunjukkan akurasi tertinggi didapatkan oleh skenario Klasifikasi MNB Label Prediksi Dengan PCA 114 sebesar 71% dengan menghasilkan data benar terprediksi sebesar 117 untuk positif dan 42 untuk negatif.

Kata kunci: *Multinomial Naïve Bayes, Seleksi Fitur, Principal Component Analysis, Analisis Sentimen, Kereta Api*

ABSTRACT

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS IN SENTIMENT ANALYSIS USING MULTINOMIAL NAÏVE BAYES

(Case Study: Public Service of Local Trains DAOP 8)

Student Name / NPM : Risnaldy Novendra Irawan / 20083010017
Study Program : Sains Data, FASILKOM, UPN Veteran Jatim
Advisor 1 : Kartika Maulida Hindrayani, S.Kom., M.Kom
Advisor 2 : Dr. Ir. Mohammad Idhom, SP., S.Kom., MT.

Abstract

This research aims to conduct sentiment analysis on the users of subsidized train services in the operational area 8 Surabaya. The method used for the analysis is the Naïve Bayes Classifier with the objective of understanding the impact of Principal Component Analysis (PCA) feature selection on the Multinomial Naïve Bayes algorithm. The data preprocessing stages included cleaning, weighting, and splitting the data, resulting in a total of 1123 data points with two classes, namely positive and negative. Subsequently, data processing was conducted to find the best PCA features and perform classification. The data processing results using Multinomial Naïve Bayes with manual labeling without PCA feature selection showed more accurate performance compared to using feature selection, achieving an accuracy of 80% during the testing process. The aforementioned scenario also resulted in correctly predicted data of 139 for positive and 40 for negative. Other results on the data showed that the highest accuracy was obtained by the scenario of MNB Classification with PCA 114, reaching 71%, with correctly predicted data of 117 for positive and 42 for negative.

Keywords: *Multinomial Naïve Bayes, Feature Selection, Principal Component Analysis, Sentiment Analysis, Train*