

**IMPLEMENTASI METODE ENSEMBLE MAJORITY VOTE  
PADA ALGORITMA NAIVE BAYES DAN RANDOM FOREST  
UNTUK ANALISIS SENTIMEN TWITTER HARGA TIKET  
PESAWAT DOMESTIK**

**SKRIPSI**



Oleh :

**DIMAS TRIYANA**

**NPM. 20081010049**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2024**

**IMPLEMENTASI METODE ENSEMBLE MAJORITY VOTE  
PADA ALGORITMA NAIVE BAYES DAN RANDOM FOREST  
UNTUK ANALISIS SENTIMEN TWITTER HARGA TIKET  
PESAWAT DOMESTIK**

**SKRIPSI**

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana  
Komputer Program Studi Informatika



Oleh :

**DIMAS TRIYANA**

**NPM. 20081010049**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2024**

## LEMBAR PENGESAHAN SKRIPSI

Judul : IMPLEMENTASI METODE ENSEMBLE MAJORITY  
VOTE PADA ALGORITMA NAIVE BAYES DAN  
RANDOM FOREST UNTUK ANALISIS SENTIMEN  
TWITTER HARGA TIKET PESAWAT DOMESTIK

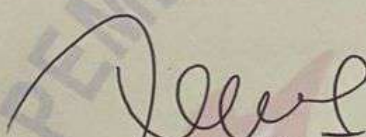
Oleh : Dimas Triyana

NPM : 20081010049

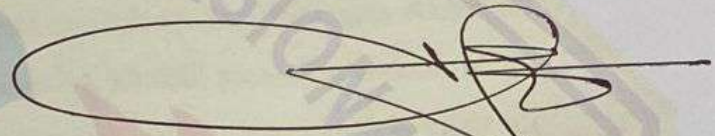
Telah Diseminarkan Dalam Ujian Skripsi Pada:  
Hari Selasa, Tanggal 21 Mei 2024

1 Dosen Pembimbing Mengetahui


1 Dosen Penguji

  
M. Muharrom Al Haromainy, S.Kom., M.Kom

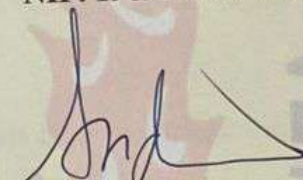
NIP. 19950601 202203 1000

  
Dr. Basuki Rahmat, S.Si. MT

NIP. 19690723 2021211 002

2  
  
Hendra Maulana, S.Kom., M.Kom

NPT. 201198 31 223248

2  
  
Andreas Nugroho Sihananto, S.Kom., M.Kom

NPT. 211199 00 412271

  
Dekan  
Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, Mt

NIP. 19681126 199403 2 001

Menyetujui

Koordinator Program Studi  
Informatika

  
Fetty Tri Anggraeny, S.Kom. M.Kom

NIP. 19820211 2021212 005

## SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Dimas Triyana

NPM : 20081010049

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

“IMPLEMENTASI METODE ENSEMBLE MAJORITY VOTE PADA ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN TWITTER HARGA TIKET PESAWAT DOMESTIK”

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan 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 Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.



Surabaya, 29 Mei 2024

Hormat saya,

  
Dimas Triyana

NPM. 20081010049



# IMPLEMENTASI METODE ENSEMBLE MAJORITY VOTE PADA ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN TWITTER HARGA TIKET PESAWAT DOMESTIK

**Nama Mahasiswa : Dimas Triyana**

**NPM : 20081010049**

**Program Studi : Informatika**

**Dosen Pembimbing : M. Muharrom Al Haromainy, S.Kom., M.Kom.**

**Hendra Maulana, S.Kom., M.Kom.**

## ABSTRAK

Transportasi sangat penting bagi pertumbuhan ekonomi dan menjadi tulang punggung pembangunan bangsa. Industri penerbangan domestik terus tumbuh pesat, dan sedang pulih setelah terdampak Covid-19. Pada 2024, diprediksi bisnis angkutan udara akan bangkit dengan lonjakan permintaan. Indonesia adalah pasar terbesar dan tercepat tumbuh di ASEAN, diperkirakan mencapai 390 juta penumpang pada 2030. Kecepatan, kenyamanan, efisiensi waktu membuat pesawat terbang lebih dipilih dibanding transportasi lain. Meski demikian, masih ada persoalan di mana harga tiket pesawat domestik menjadi salah satu faktor krusial yang memengaruhi keputusan pembelian konsumen. Dengan meningkatnya jumlah penumpang pesawat domestik mengakibatkan opini yang bervariasi terkait harga tiket pesawat domestik pada media sosial. Analisis sentimen, menggunakan teknologi pengolahan, adalah pendekatan efektif untuk memahami opini publik tentang harga tiket pesawat. Ini bertujuan untuk memberikan informasi kepada pemerintah dan pemangku kepentingan lainnya mengenai opini masyarakat untuk pengambilan keputusan yang lebih bijak.

Fokus penelitian melibatkan analisis sentimen twitter menggunakan metode *Naive Bayes*, *Random Forest* dan *Ensemble Majority Voting*, dengan hasil perbandingan antara algoritma individu tersebut yang menunjukkan performa terbaik pada tingkat akurasi dan apakah efektif mengimplementasi metode *ensemble majority vote* dalam meningkatkan akurasi analisis sentimen.

Hasil penelitian mencatat bahwa pada skenario pertama *Naive Bayes*

menggunakan data *training* dan *testing* dengan rasio 80:20, menghasilkan akurasi sebesar 70,31%. Skenario kedua, dengan rasio 70:30, mendapat akurasi 71,75%. Metode *Random Forest* juga digunakan sebagai algoritma perbandingan. Pada skenario pertama, *Random Forest* mencapai akurasi 84,91%, dan pada skenario kedua, 85,71%. Penggabungan kedua algoritma menggunakan *ensemble majority voting* menghasilkan akurasi 85,88% dan 85,87% untuk kedua skenario, menunjukkan peningkatan akurasi prediksi dibandingkan model individu algoritma. Dalam artian metode *ensemble majority vote* dinilai efektif dalam meningkatkan akurasi analisis sentimen terhadap harga tiket pesawat domestik. Arah opini masyarakat terkait harga tiket pesawat domestik dalam media sosial Twitter adalah negatif.

**Kata Kunci:** Analisis Sentimen, *Twitter*, *Naive Bayes*, *Random Forest*, *Ensemble Majority Vote*, Harga, Tiket, Pesawat Domestik

## KATA PENGANTAR

Dengan penuh rasa syukur, peneliti ingin mengucapkan terima kasih kepada Tuhan YME atas berkat dan rahmat-Nya yang telah memungkinkan penyelesaian skripsi berjudul " Implementasi Metode Ensemble Majority Vote Pada Algoritma Naive Bayes Dan Random Forest Untuk Analisis Sentimen Twitter Harga Tiket Pesawat Domestik".

Laporan skripsi ini disusun dalam rangka memenuhi persyaratan mata kuliah skripsi dan sebagai salah satu langkah penting dalam menyelesaikan Pendidikan tingkat Sarjana (S1) di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Peneliti sadar akan adanya keterbatasan dalam penulisan skripsi ini, dan dengan rendah hati menerima saran serta kritik yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, baik bagi pembaca maupun untuk penelitian mendatang.

Surabaya, 29 Mei 2024

**Dimas Triyana**  
**NPM. 20081010049**

## UCAP TERIMA KASIH

Puji dan syukur peneliti panjatkan kehadiran 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. Kedua orang tua serta keluarga yang selalu memberikan dukungan moral dan spiritual yang tiada henti, yang menjadi sumber semangat dan inspirasi penulis untuk terus berjuang menyelesaikan skripsi ini.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT, selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Ir. Novirina Hendrasarie, S.T, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak M. Muharrom Al Haromainy, 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.
6. Bapak Hendra Maulana, 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.
7. Dosen Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang membantu dan memberi masukan dalam segala hal dalam pengerjaan skripsi ini.
8. Seluruh teman-teman yang telah memberikan dukungan, saran, bantuan dan semangat hingga penulis dapat menyelesaikan penulisan skripsi ini.

Akhir kata, semoga Allah SWT senantiasa membalas semua kebaikan yang telah diberikan. Semoga hasil dari penelitian ini dapat bermanfaat bagi para pembaca serta memberikan pemikiran baru yang bermanfaat bagi pihak yang membutuhkan.



## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
SURAT PERNYATAAN BEBAS DARI PLAGIASI.....	ii
ABSTRAK .....	iii
KATA PENGANTAR.....	v
UCAP TERIMA KASIH.....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	x
DAFTAR TABLE .....	xii
DAFTAR KODE PROGRAM .....	xiii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan .....	5
1.4. Manfaat .....	5
1.5. Batasan Masalah .....	6
BAB II TINJUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu .....	7
2.2. Harga Tiket Pesawat Domestik.....	10
2.3. Twitter.....	10
2.4. Analisis Sentimen .....	11
2.5. <i>Data Mining</i> .....	11
2.6. <i>Preprocessing</i> .....	11
2.6.1. <i>Cleaning</i> .....	13
2.6.2. <i>Case-Folding</i> .....	13

2.6.3. <i>Spelling Normalization</i> .....	13
2.6.4. <i>Tokenization</i> .....	13
2.6.5. <i>Stopword Removal</i> .....	13
2.6.6. <i>Stemming</i> .....	14
2.7. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	13
2.8. <i>Naïve Bayes</i> .....	15
2.9. <i>Random Forest</i> .....	16
2.10. <i>Ensemble Majority Voting</i> .....	17
2.11. <i>Confusion Matrix</i> .....	18
<b>BAB III METODOLOGI</b> .....	<b>21</b>
3.1. Tahap Penelitian .....	21
3.2. Pengumpulan Data.....	22
3.3. <i>Preprocessing</i> .....	22
3.3.1. <i>Case-Folding</i> .....	23
3.3.2. <i>Cleaning</i> .....	23
3.3.3. <i>Tokenization</i> .....	24
3.3.4. <i>Stopword Removal</i> .....	25
3.3.5. <i>Spelling Normalization</i> .....	26
3.3.6. <i>Stemming</i> .....	27
3.4. <i>Labelling Data</i> .....	27
3.5. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	28
3.6. Skema Pengujian.....	31
3.7. <i>Naïve Bayes</i> .....	31
3.8. <i>Random Forest</i> .....	35
3.9. <i>Ensemble Majority Voting</i> .....	40
3.10. <i>Confusion Matrix</i> .....	43
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>45</b>
4.1. Pengumpulan Data.....	45

4.2. <i>Preprocessing</i> .....	46
4.3. Labeling Data.....	58
4.4. Pembobotan Kata (TF-IDF).....	61
4.5. Skenario Pengujian .....	68
4.6. Klasifikasi Naive Bayes.....	69
4.7. Klasifikasi Random Forest.....	77
4.8. Ensemble Majority Voting.....	79
4.9. Hasil dan Analisa .....	82
BAB V KESIMPULAN DAN SARAN.....	87
5.1. Kesimpulan .....	87
5.2. Saran .....	87
DAFTAR PUSTAKA.....	89

## DAFTAR GAMBAR

Gambar 2.1 Algoritma sederhana Random Forest .....	16
Gambar 3. 1 Tahap Penelitian .....	21
Gambar 3. 2 Alur <i>Labelling</i> .....	28
Gambar 3. 3 Diagram alir <i>TF-IDF</i> .....	29
Gambar 3. 4 Hasil TF-IDF .....	31
Gambar 3. 5 Diagram alir <i>Naive Bayes</i> .....	32
Gambar 3. 6 Diagram alir <i>Random Forest</i> .....	36
Gambar 3. 7 Frekuensi Kemunculan Kata (Random Forest).....	38
Gambar 3. 8 Alur <i>Majority Voting</i> .....	41
Gambar 4. 1 Dataset Hasil <i>Crawling</i> .....	46
Gambar 4. 2 Pembatasan pada dataset yang digunakan.....	47
Gambar 4. 3 Hasil <i>Case Folding</i> .....	48
Gambar 4. 4 Hasil <i>Cleaning Data</i> .....	50
Gambar 4. 5 Hasil <i>Tokenizing</i> .....	51
Gambar 4. 6 Hasil <i>Frequency Tokens</i> .....	52
Gambar 4. 7 <i>List File Stopword</i> .....	53
Gambar 4. 8 <i>Stopword Removal</i> .....	54
Gambar 4. 9 <i>List File External</i> Normalisasi .....	55
Gambar 4. 10 Hasil Normalisasi .....	56
Gambar 4. 11 Hasil <i>Stemming</i> .....	58
Gambar 4. 12 Hasil Translate.....	59
Gambar 4. 13 Hasil <i>Labeling</i> data.....	60
Gambar 4. 14 Status <i>Labeling</i> Data .....	61
Gambar 4. 15 Dataset pembobotan kata .....	62
Gambar 4. 16 <i>List</i> Dataset.....	63
Gambar 4. 17 Perhitungan <i>Term Frequency</i> .....	64
Gambar 4. 18 Hasil <i>Term Frequency</i> .....	65
Gambar 4. 19 Hasil TF_IDF .....	68
Gambar 4. 20 Hasil Akurasi Pengujian pertama NB .....	71
Gambar 4. 21 Hasil Akurasi Pengujian pertama NB kelas positif dan negatif.....	72
Gambar 4. 22 Hasil Sentimen Skenario Pertama <i>Naive Bayes</i> .....	73

Gambar 4. 23 Visualisasi <i>Classification Report</i> Skenario 1 .....	73
Gambar 4. 24 Visualisasi Confusion Matrix skenario pertama .....	74
Gambar 4. 25 Hasil Akurasi Pengujian kedua NB.....	74
Gambar 4. 26 Hasil Akurasi Pengujian kedua NB kelas positif dan negatif .....	74
Gambar 4. 27 Hasil <i>Classification Report</i> Skenario Kedua <i>Naive Bayes</i> .....	75
Gambar 4. 28 Sentimen Skenario Kedua <i>Naive Bayes</i> .....	75
Gambar 4. 29 Visualisasi <i>Confusion Matrix</i> kedua .....	76
Gambar 4. 30 Visualisasi word clouds.....	76
Gambar 4. 31 Hasil Akurasi dan <i>Classification Report</i> .....	77
Gambar 4. 32 Hasil Confusion Matrix.....	78
Gambar 4. 33 Hasil Akurasi dan <i>Classification Report</i> .....	79
Gambar 4. 34 Hasil <i>Confusion Matrix</i> data kedua.....	79
Gambar 4. 35 <i>Classification Report</i> skenario pertama <i>Ensemble Majority Voting</i> .	80
Gambar 4. 36 Hasil <i>Confusion Matrix</i> .....	80
Gambar 4. 37 <i>Classification Report</i> skenario kedua <i>Ensemble Majority Voting</i> ..	81
Gambar 4. 38 Hasil <i>Confusion Matrix</i> Skenario Kedua .....	81
Gambar 4. 39 Visualisasi Kata <i>Majority Voting</i> .....	82

## DAFTAR TABLE

Table 2.1 Penelitian Terdahulu .....	7
Table 2. 2 <i>Confusion Matrix</i> .....	18
Table 3. 1 Proses <i>Case-folding</i> .....	23
Table 3. 2 Proses <i>Cleaning</i> .....	24
Table 3. 3 Proses <i>Tokenization</i> .....	24
Table 3. 4 Proses <i>Stopword Removal</i> .....	25
Table 3. 5 Proses <i>Spelling Normalization</i> .....	26
Table 3. 6 Proses <i>Stemming</i> .....	27
Table 3. 7 Hasil <i>Labelling</i> .....	28
Table 3. 8 Dokumen.....	29
Table 3. 9 Table Uji Coba Model.....	32
Table 3. 10 Contoh Data Ulasan Untuk <i>Random Forest</i> .....	36
Table 3. 11 Corpus Ulasan (Random Forest).....	37
Table 3. 12 Kumpulan Prediksi dari Setiap Model .....	41
Table 3. 13 Hasil <i>Majority Voting</i> .....	43
Table 4. 1 Skema Pengujian.....	69
Table 4. 2 <i>Confussion Matrix Naive Bayes</i> .....	83
Table 4. 3 <i>Confussion Matrix Random Forest</i> .....	83
Table 4. 4 <i>Confussion Matrix Random Forest</i> .....	84
Table 4. 5 Perbandingan Akurasi .....	85

## DAFTAR KODE PROGRAM

Kode Program 4. 1 Pembatasan Dataset .....	46
Kode Program 4. 2 <i>Case Folding</i> .....	47
Kode Program 4. 3 <i>Cleaning Data</i> .....	49
Kode Program 4. 4 <i>Tokenizing</i> .....	51
Kode Program 4. 5 Frekuensi <i>Tokenizing</i> .....	51
Kode Program 4. 6 <i>Stopword Removal</i> .....	53
Kode Program 4. 7 Normalisasi Kata .....	55
Kode Program 4. 8 <i>Stemmer</i> .....	57
Kode Program 4. 9 Terjemahan .....	59
Kode Program 4. 10 <i>Labeling data</i> .....	60
Kode Program 4. 11 <i>Prepare Data</i> .....	62
Kode Program 4. 12 <i>List Kata</i> .....	63
Kode Program 4. 13 <i>Term Frequency</i> .....	64
Kode Program 4. 14 Output <i>Term Frequency</i> .....	65
Kode Program 4. 15 Perhitungan <i>Document Frequency (DF)</i> .....	66
Kode Program 4. 16 Perhitungan <i>IDF</i> .....	66
Kode Program 4. 17 Perhitungan <i>TF-IDF</i> .....	67
Kode Program 4. 18 Output <i>TF-IDF</i> .....	68
Kode Program 4. 19 Pembagian Data .....	69
Kode Program 4. 20 Skenario 1 Klasifikasi Naive Bayes .....	70
Kode Program 4. 21 Akurasi Negatif dan Positif Skenario 1 .....	71
Kode Program 4. 22 Sentimen Skenario Pertama Labelling.....	72
Kode Program 4. 23 <i>Classification Report</i> dan <i>Confusion Matrix</i> .....	73
Kode Program 4. 24 Sentimen Dengan <i>Majority Voting</i> .....	82