

**SCR - Data Mining**

**Analisis Sentimen Kenaikan Pajak Hiburan Pada X Menggunakan  
Majority Vote Pada Metode Naive Bayes dan Support Vector Machine**

**SKRIPSI**



Oleh:

**BAYU SAMODERA**

**20081010048**

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



## LEMBAR PENGESAHAN SKRIPSI

**Judul** : ANALISIS SENTIMEN KENAIKAN PAJAK HIBURAN  
PADA X MENGGUNAKAN MAJORITY VOTE PADA  
METODE NAIVE BAYES DAN SUPPORT VECTOR  
MACHINE

**Oleh** : Bayu Samodera

**NPM** : 20081010048

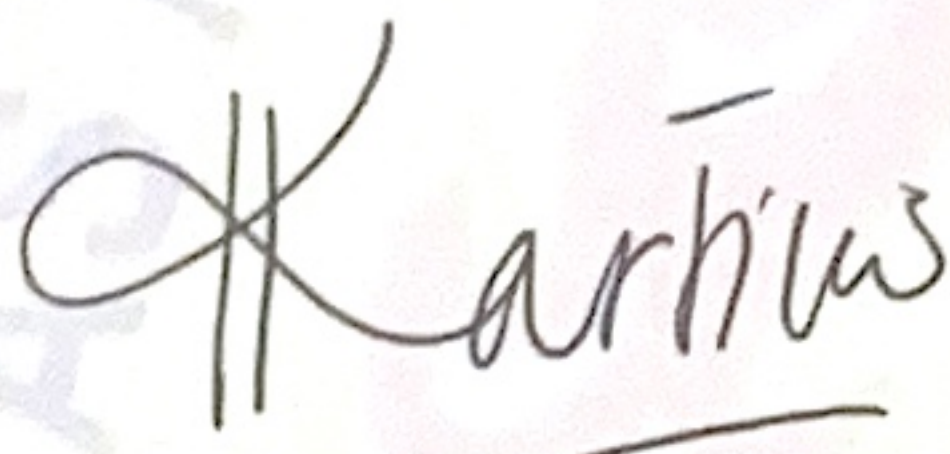
Telah Diseminarkan Dalam Ujian Skripsi Pada :

Hari Kamis, Tanggal 4 Juli 2024

### Mengetahui

#### Dosen Pembimbing

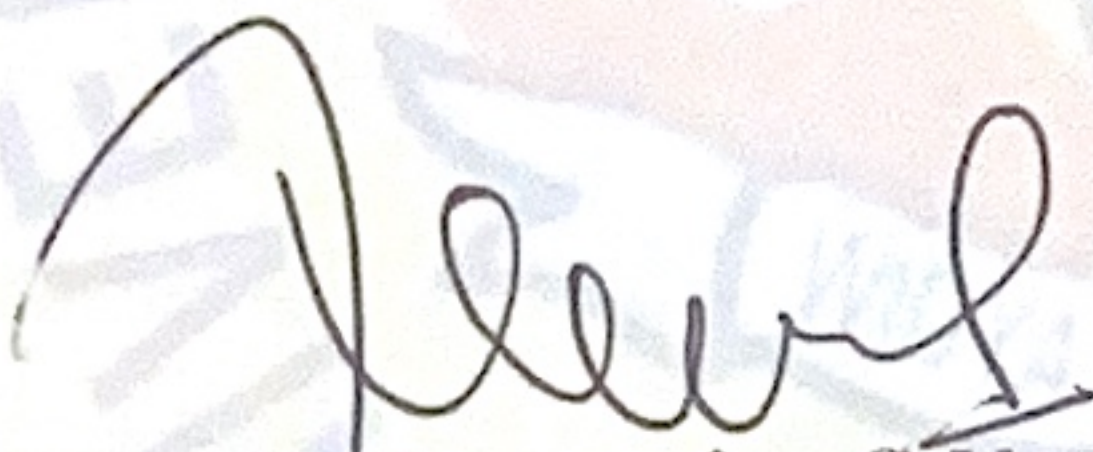
1.



Dr. Ir. Kartini, S.Kom. MT.

NIP : 19611110 199103 2 001

2.

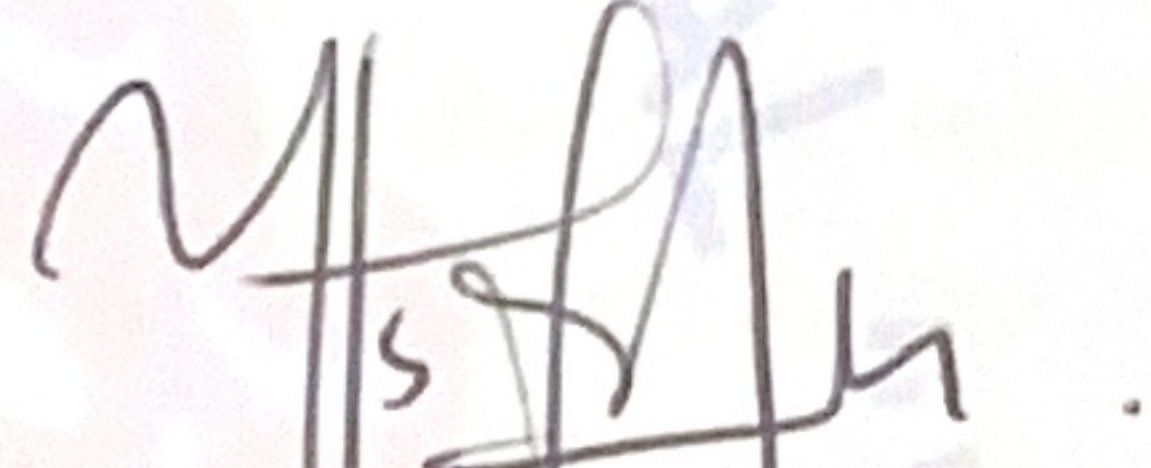


Muharrom Al Haromainy, S.Kom., M.Kom.

NIP : 19950601 202203 1 006

#### Dosen Penguji

1.



Yisti Vita Via, S.ST. M.Kom.

NIP : 19860425 202121 2 001

2.



Firza Prima Aditiawan, S.Kom., MTI

NIP : 19860523 202121 1 003

### Menyetujui

#### Dekan Fakultas Ilmu

#### Komputer

  
Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP : 19681126 199403 2 001

#### 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 : Bayu Samodera

NPM : 20081010048

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

**"ANALISIS SENTIMEN KENAIKAN PAJAK HIBURAN PADA  
MENGUNAKAN MAJORITY VOTE PADA METODE NAIVE BAYES DAN  
SUPPORT VECTOR MACHINE"**

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, 14 Juli 2024

Hormat saya,



Bayu Samodera

NPM. 20081010048





# ANALISIS SENTIMEN KENAIKAN PAJAK HIBURAN PADA MENGGUNAKAN MAJORITY VOTE PADA METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE

**Nama Mahasiswa : Bayu Samodera**

**NPM : 20081010048**

**Program Studi : Teknik Informatika**

**Dosen Pembimbing : Dr. Ir. Kartini, S.Kom. MT.**

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

## ABSTRAK

Pajak hiburan adalah pajak yang dikenakan pada kegiatan atau layanan hiburan seperti bioskop, konser, atau lainnya. Kenaikan tarif pajak hiburan dapat berdampak signifikan terhadap industri hiburan dan persepsi masyarakat terhadap kebijakan pemerintah. Untuk itu, sangat penting melaksanakan analisis sentimen terkait kenaikan pajak hiburan guna memahami respon dan sikap masyarakat.

Penelitian ini menggunakan data dari media sosial X untuk mengumpulkan opini masyarakat terkait kenaikan pajak hiburan. Metode *Support Vector Machine Naive Bayes* diterapkan secara individual dan kemudian diintegrasikan melalui metode *Majority Vote* untuk mengklasifikasikan sentimen sebagai positif, negatif, atau netral.

Pada penelitian ini, data yang digunakan berjumlah 4012 *tweet*. Penelitian dilakukan dengan tiga skenario. Pada skenario pertama (90% data latih, 10% data uji), *Naive Bayes* mencapai akurasi 71.82%, *Support Vector Machine* 77.56, dan *Majority Vote* 80.35%. Pada skenario kedua (80% data latih, 20% data uji), *Naive Bayes* mencapai akurasi 70.32%, *Support Vector Machine* 75.06%, dan *Majority Vote* 79.08%. Pada skenario ketiga (70% data latih, 30% data uji), *Naive Bayes* mencapai akurasi 68.16%, *Support Vector Machine* 72.82%, dan *Majority Vote* 76.25%. Hal ini menunjukkan bahwa penggabungan *Naive Bayes* dan *Support Vector Machine* melalui teknik *Majority Vote* dapat meningkatkan akurasi dalam klasifikasi sentimen.

**Kata kunci :** Analisis, *Naive Bayes*, *Support Vector Machine*, *Majority Vote*.

## **KATA PENGANTAR**

Dengan segala kerendahan hati, penulis ingin mengucapkan puji syukur ke hadirat Allah SWT Yang Maha Pengasih dan Maha Penyayang, karena atas segala rahmat, ridho, dan karunia-Nya, penulis dapat menyelesaikan tugas akhir dengan judul “ANALISIS SENTIMEN KENAIKAN PAJAK HIBURAN PADA MENGGUNAKAN MAJORITY VOTE PADA METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE”.

Penulis juga ingin menyampaikan terimakasih yang tulus kepada semua pihak yang telah memberikan dukungan, bantuan, dan bimbingan selama penelitian ini berlangsung.

Penulis sadar bahwa laporan ini masih memiliki kekurangan karena keterbatasan pengetahuan dan kemampuan penulis. Oleh karena itu, penulis sangat mengharapkan masukan, kritik, dan saran yang membangun untuk memperbaiki laporan ini.

Surabaya, 14 Juli 2024

Penulis

## UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Allah SWT karena atas segala berkah dan karunia-Nya penulis dapat menyelesaikan tugas akhir ini dengan baik.

Penulis sadari bahwa dalam menyelesaikan laporan akhir ini banyak pihak yang telah membantu memberi bimbingan, arahan, dan doa yang akan selalu penulis kenang dan syukuri. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada :

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. 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 Dr. Ir. Kartini, S.Kom. MT selaku dosen pembimbing 1 dan bapak M. Muharrom Al Haromany, S.Kom., M.Kom selaku dosen pembimbing 2 yang telah membimbing serta memberikan arahan dalam menyusun pengerjaan tugas akhir ini.
5. Orang Tua tercinta serta kakak saya yang telah memberikan doa, dukungan, dan motivasi yang diberikan kepada penulis.
6. Seluruh teman seangkatan yang telah saling mendukung dan berbagi pengalaman selama proses penulisan skripsi ini.

Penulis menyadari bahwa dalam penulisan laporan ini masih banyak terdapat kekurangan, untuk itu penulis sangat mengharapkan saran dan kritik yang bersifat membangun dalam penyempurnaan penulisan laporan ini. Akhirnya, dengan mengharap Ridho dari Allah SWT, semoga laporan tugas akhir ini bermanfaat bagi kita semua. Aamiin.

Surabaya, 14 Juli 2024

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
SURAT PERNYATAAN BEBAS DARI PLAGIASI.....	iii
ABSTRAK .....	iv
KATA PENGANTAR .....	v
UCAPAN TERIMAKASIH.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xii
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.....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1 Penelitian Terdahulu .....	7
2.2 Pajak Hiburan .....	10
2.3 Analisis Sentimen.....	10
2.4 <i>Twitter (X)</i> .....	10
2.5 <i>Tweet Harvest</i> .....	11
2.6 <i>Text Mining</i> .....	11
2.7 <i>Text Preprocessing</i> .....	12
2.8 TF-IDF.....	13

2.9	Klasifikasi.....	14
2.10	<i>Naive Bayes</i> .....	15
2.11	<i>Support Vector Machine</i> .....	17
2.12	<i>Majority Voting</i> .....	20
2.13	<i>Confusion Matrix</i> .....	21
BAB III METODOLOGI .....		23
3.1	Studi Literatur .....	23
3.2	Analisis Kebutuhan .....	24
3.2.1	Kebutuhan Data.....	24
3.2.2	Kebutuhan <i>Hardware</i> dan <i>Software</i> .....	24
3.3	Perancangan Model.....	24
3.3.1	Pengumpulan Data .....	25
3.3.2	Penyaringan Data .....	26
3.3.3	<i>Pre-processing</i> Data .....	26
3.3.4	Pelabelan Data.....	31
3.3.5	Pembobotan Data TF-IDF .....	31
3.3.6	Klasifikasi <i>Naive Bayes</i> .....	33
3.3.7	Klasifikasi <i>Support Vector Machine</i> .....	37
3.3.8	<i>Majority Voting</i> .....	42
3.3.9	Evaluasi Akhir ( <i>Confusion Matrix</i> ).....	44
3.4	Visualisasi .....	46
BAB IV HASIL DAN PEMBAHASAN .....		47
4.1	Pengumpulan Data .....	47
4.2	Penyaringan Data .....	48
4.3	<i>Pre-processing</i> Data.....	48



4.4	Pelabelan Data.....	58
4.5	Ekstrasi Fitur TF-IDF.....	59
4.6	Klasifikasi <i>Naive Bayes</i> .....	63
4.7	Klasifikasi <i>Support Vector Machine</i> .....	69
4.8	<i>Majority Voting</i> .....	75
4.9	Tabel Perbandingan.....	81
4.10	Visualisasi .....	82
BAB V KESIMPULAN DAN SARAN .....		84
5.1	Kesimpulan .....	84
5.2	Saran.....	85
DAFTAR PUSTAKA .....		86



## DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i> .....	21
Tabel 3. 1 Contoh Dataset <i>Case Folding</i> .....	27
Tabel 3. 2 Contoh Dataset <i>Cleaning</i> .....	28
Tabel 3. 3 Contoh Dataset <i>Tokenizing</i> .....	28
Tabel 3. 4 Contoh Dataset Normalisasi .....	29
Tabel 3. 5 Contoh Dataset <i>Stopword Removal</i> .....	30
Tabel 3. 6 Contoh Dataset <i>Stemming</i> .....	30
Tabel 3. 7 Contoh Dataset Hasil Pelabelan.....	31
Tabel 3. 8 Contoh Dataset.....	32
Tabel 3. 9 Hasil perhitungan TF-IDF pada tiap <i>term</i> .....	33
Tabel 3. 10 Contoh data latih dan data uji .....	34
Tabel 3. 11 Hasil perhitungan probabilitas .....	36
Tabel 3. 12 Contoh dataset kelas sentimen.....	38
Tabel 3. 13 Hasil perhitungan matriks fitur.....	39
Tabel 3. 14 Hasil perhitungan matriks <i>Hessian</i> .....	39
Tabel 3. 15 Hasil perhitungan nilai $E_i$ .....	40
Tabel 3. 16 Hasil perhitungan nilai $\alpha_i$ .....	40
Tabel 3. 17 Tabel <i>Confusion Matrix</i> .....	44
Tabel 4. 1 <i>Source Code</i> Pengumpulan Data .....	47
Tabel 4. 2 <i>Source Code Case Folding</i> .....	48
Tabel 4. 3 <i>Source Code</i> Menghilangkan <i>URL</i> .....	49
Tabel 4. 4 <i>Source Code</i> Menghilangkan <i>Mention</i> .....	50
Tabel 4. 5 <i>Source Code</i> Menghilangkan <i>Hastag</i> .....	51
Tabel 4. 6 <i>Source Code</i> Menghilangkan Nomor/Angka.....	51
Tabel 4. 7 <i>Source Code</i> Menghilangkan Tanda Baca.....	52
Tabel 4. 8 <i>Source Code</i> Menghilangkan Simbol .....	53
Tabel 4. 9 <i>Source Code</i> Menghilangkan Spasi Lebih.....	53
Tabel 4. 10 <i>Source Code Tokenizing</i> .....	54
Tabel 4. 11 <i>Source Code</i> Normalisasi .....	55
Tabel 4. 12 <i>Source Code Stopword Removal</i> .....	56
Tabel 4. 13 <i>Source Code Stemming</i> .....	57



Tabel 4. 14 <i>Source Code</i> Pelabelan Data.....	58
Tabel 4. 15 <i>Source Code Term Frequency</i> .....	60
Tabel 4. 16 <i>Source Code Inverse Document Frequency</i> .....	61
Tabel 4. 17 <i>Source Code TF-IDF</i> .....	62
Tabel 4. 18 <i>Source Code Multinomial Naive Bayes</i> .....	63
Tabel 4. 19 <i>Source Code SVM kernel Polynomial</i> .....	69
Tabel 4. 20 <i>Source Code Majority Voting</i> .....	75
Tabel 4. 21 <i>Source Code Sentimen Akhir</i> .....	80
Tabel 4. 22 Perbandingan Akurasi.....	81
Tabel 4. 23 <i>Source Code Word Cloud</i> .....	82



## DAFTAR GAMBAR

Gambar 2. 1 Tahapan <i>Text Preprocessing</i> .....	12
Gambar 2. 2 <i>Hyperplane</i> pada SVM .....	17
Gambar 3. 1 Alur Metodologi Penelitian .....	23
Gambar 3. 2 Diagram Alir Perancangan Model Penelitian .....	25
Gambar 3. 3 Diagram Alir <i>Pre-processing</i> Data .....	26
Gambar 3. 4 Diagram Alir <i>Naive Bayes</i> .....	34
Gambar 3. 5 Diagram Alir <i>Support Vector Machine</i> .....	37
Gambar 3. 6 Diagram Alir <i>Majority Vote</i> .....	42
Gambar 3. 7 <i>Headmap Matrix</i> .....	46
Gambar 3. 8 <i>Word Cloud</i> Sentimen Negatif.....	46
Gambar 4. 1 Hasil Pengumpulan Data Dalam File .csv .....	47
Gambar 4. 2 Hasil Penyaringan Data.....	48
Gambar 4. 3 Hasil <i>Case Folding</i> .....	49
Gambar 4. 4 Hasil Menghilangkan <i>URL</i> .....	50
Gambar 4. 5 Hasil Menghilangkan <i>Mention</i> .....	50
Gambar 4. 6 Hasil Menghilangkan <i>Hastag</i> .....	51
Gambar 4. 7 Hasil Menghilangkan nomor/angka.....	52
Gambar 4. 8 Hasil Menghilangkan Tanda Baca.....	52
Gambar 4. 9 Hasil Menghilangkan Simbol .....	53
Gambar 4. 10 Hasil Menghilangkan Spasi Lebih.....	54
Gambar 4. 11 Hasil <i>Tokenizing</i> .....	55
Gambar 4. 12 Hasil <i>Normalisasi</i> .....	56
Gambar 4. 13 Hasil <i>Stopword Removal</i> .....	57
Gambar 4. 14 Hasil <i>Stemming</i> .....	58
Gambar 4. 15 Hasil Pelabelan Data .....	59
Gambar 4. 16 Grafik Hasil Pelabelan Data.....	59
Gambar 4. 17 Output Hasil <i>Term Frequency</i> Pada Tiap Kata.....	60
Gambar 4. 18 Output Hasil <i>Inverse Document Frequency</i> .....	62
Gambar 4. 19 Output Hasil TF-IDF .....	63
Gambar 4. 20 Output Hasil Skenario 1 (90% data <i>train</i> ) .....	65
Gambar 4. 21 Hasil Akurasi Setiap Label Skenario 1 .....	65
Gambar 4. 22 Hasil <i>Classification Report</i> Skenario 1 .....	65



Gambar 4. 23 Hasil <i>Confusion Matrix</i> Skenario 1.....	66
Gambar 4. 24 Output Hasil Skenario 2 (80% data <i>train</i> ) .....	66
Gambar 4. 25 Hasil Akurasi Setiap Label Skenario 2 .....	66
Gambar 4. 26 Hasil <i>Classification Report</i> Skenario 2.....	67
Gambar 4. 27 Hasil <i>Confusion Matrix</i> Skenario 2.....	67
Gambar 4. 28 Output Hasil Skenario 3 (70% data <i>train</i> ) .....	68
Gambar 4. 29 Hasil Akurasi Setiap Label Skenario 3 .....	68
Gambar 4. 30 Hasil <i>Classification Report</i> Skenario 3.....	68
Gambar 4. 31 Hasil <i>Confusion Matrix</i> Skenario 3.....	69
Gambar 4. 32 Output Hasil Skenario 1 SVM (90% data <i>train</i> ).....	71
Gambar 4. 33 Hasil Akurasi Setiap Label Skenario 1 .....	71
Gambar 4. 34 Hasil <i>Classification Report</i> Skenario 1 .....	71
Gambar 4. 35 Hasil <i>Confusion Matrix</i> Skenario 1.....	72
Gambar 4. 36 Output Hasil Skenario 2 SVM (80% data <i>train</i> ).....	72
Gambar 4. 37 Hasil Akurasi Setiap Label Skenario 2 .....	72
Gambar 4. 38 Hasil <i>Classification Report</i> Skenario 2.....	73
Gambar 4. 39 Hasil <i>Confusion Matrix</i> Skenario 2.....	73
Gambar 4. 40 Output Hasil Skenario 3 SVM (70% data <i>train</i> ) .....	74
Gambar 4. 41 Hasil Akurasi Setiap Label Skenario 3 .....	74
Gambar 4. 42 Hasil <i>Classification Report</i> Skenario 3.....	74
Gambar 4. 43 Hasil <i>Confusion Matrix</i> Skenario 3.....	75
Gambar 4. 44 Hasil Akurasi Skenario 1 (90% data <i>train</i> ) .....	76
Gambar 4. 45 Hasil <i>Classification Report Majority Vote</i> Skenario 1 .....	76
Gambar 4. 46 Hasil <i>Confusion Matrix Majority Vote</i> Skenario 1 .....	77
Gambar 4. 47 Hasil Akurasi Skenario 2 (80% data <i>train</i> ).....	77
Gambar 4. 48 Hasil <i>Classification Report Majority Vote</i> Skenario 2 .....	78
Gambar 4. 49 Hasil <i>Confusion Matrix Majority Vote</i> Skenario 2 .....	78
Gambar 4. 50 Hasil Akurasi Skenario 3 (70% data <i>train</i> ).....	79
Gambar 4. 51 Hasil <i>Classification Report Majority Vote</i> Skenario 3 .....	79
Gambar 4. 52 Hasil <i>Confusion Matrix Majority Vote</i> Skenario 3 .....	79
Gambar 4. 53 Hasil Akhir Sentimen <i>Majority Voting</i> .....	81
Gambar 4. 54 Word Cloud Sentimen Positif .....	83
Gambar 4. 55 Word Cloud Sentimen Negatif.....	83
Gambar 4. 56 Word Cloud Sentimen Netral.....	83