

**ANALISIS SENTIMEN PEMINDAHAN IBU KOTA INDONESIA PADA
TWITTER DENGAN MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

Diajukan Untuk Memenuhi Persyaratan
Dalam Memperoleh Gelar Sarjana Komputer
Program Studi Sistem Informasi



Disusun Oleh:

Erica Mas'udah
1635010004

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

SURABAYA

2020

SKRIPSI

ANALISIS SENTIMEN PEMINDAHAN IBU KOTA INDONESIA PADA TWITTER DENGAN MENGGUNAKAN METODE NAÏVE BAYES

Disusun Oleh :


ERICA MAS'UDAH
NPM. 1635010004

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Pada Tanggal 29 April 2020

Menyetujui

Dosen Pembimbing :

1.



Eka Dyar Wahyuni, S.Kom., M.Kom
NPT. 3 8412 13 0356 1

2.



Amalia Anjam Arifiyanti, S.Kom., M.Kom
NIP. 19920812 201803 2 001

Dosen Penguji :

1.



Agung Brastama Putra, S.Kom., M.Kom
NPT. 3 8511 13 0357 1

2.



Tri Lathif Mardi Sanyanto, S.Kom., M.T
NPT. 3 8902 13 0352 1

3.



Eristya Maya Safitri, S.Kom., M.Kom
NIP. 19930316 20190320 20

Mengetahui,
Dekan Fakultas Ilmu Komputer
Universitas Pembangunan Nasional "Veteran" Jawa Timur



Dr. Ir. N. Ketut Sari, MT
NPT. 19670731 199203 2 001



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

Jl. Rungkur Madya Gunung Anyar Surabaya 60294
Telp (031) 8706369, 8783189 Fax (031) 8706372 Website www.upnjatim.ac.id

SURAT PERNYATAAN

Saya, Mahasiswa Sistem Informasi Universitas Pembangunan Nasional "Veteran" Jawa Timur, yang bertanda tangan di bawah ini :

Nama : Erica Mas'udah

NPM : 1635010004

Program Studi : Sistem Informasi

Fakultas : Ilmu Komputer

Menyatakan Bahwa Judul Skripsi / Tugas Akhir Saya Sebagai Berikut:

**ANALISIS SENTIMEN PEMINDAHAN IBU KOTA INDONESIA PADA
TWITTER DENGAN MENGGUNAKAN METODE NAÏVE BAYES**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian Orang Lain dan juga bukan merupakan Produk / Perangkat Lunak / Hasil Karya yang Saya beli dari pihak 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 institusi Pendidikan Lain.

Jika ternyata dikemudian hari Pernyataan ini terbukti Tidak Benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensinya, termasuk Pembatalan Ijazah di kemudian hari.

Hormat Saya,

Erica Mas'udah

KATA PENGANTAR

Alhamdulillah *robbil 'alamin*, segala puji bagi Allah Yang Maha Kuasa yang telah memberikan kekuatan-Nya, serta sholawat dan salam tercurahkan pada junjungan Nabi Besar Muhammad SAW, sehingga skripsi berjudul Analisis Sentimen Pemandangan Ibu Kota Indonesia pada Twitter dengan Menggunakan Metode Naïve Bayes yang merupakan persyaratan dalam menyelesaikan Program Studi Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur ini dapat diselesaikan tepat waktu.

Melalui skripsi ini, saya merasa mendapat kesempatan besar untuk lebih memperdalam ilmu pengetahuan yang diperoleh selama di perkuliahan, terutama berkenaan dengan implementasi sistem informasi dalam kehidupan sehari-hari. Namun demikian, saya menyadari bahwa skripsi ini masih memiliki banyak kelemahan dan kekurangan.

Secara khusus, dalam kesempatan ini pula, ingin mengucapkan terima kasih dan penghargaan sebesar-besarnya kepada:

1. Mama Ernie Cahyaningati selaku ibu saya yang tidak ada hentinya untuk memberikan dukungan dan kasih sayangnya kepada saya.
2. Alm. Abah Mansur Anwar yang telah berada di surga. Terima kasih untuk 19 tahun yang berharga telah mengajarkan berbagai hal untuk saya.
3. Ibu Eka Dyar Wahyuni, S.Kom, M.Kom selaku dosen pembimbing I yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi.
4. Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom selaku dosen pembimbing II yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi.

5. Seluruh dosen Program Studi Sistem Informasi atas ilmu-ilmu yang di berikan selama masa perkuliahan.
6. Teman-teman KIMS (Jinjin, Fariska, Cuny) dan Afifatul yang selama ini selalu menemani saya dalam suka dan duka.
7. Ega Febri Dharmawan, S. Kom selaku alumni pembimbing.
8. Teman-teman Program Studi Sistem Informasi yang telah membantu saya dalam menyelesaikan skripsi ini.

Terima kasih kepada semua pihak yang telah membantu penyelesaian skripsi ini. Semoga Allah membalas dengan balasan sebaik-baiknya. Dalam pengerjaan skripsi ini saya menyadari bahwa skripsi ini masih ada kekurangan. Sebagai manusia biasa pasti mempunyai keterbatasan dan banyak sekali kekurangan, terutama dalam pembuatan skripsi ini. Untuk itu saya sangat membutuhkan kritik dan saran yang membangun dalam memperbaiki penulisan skripsi ini.

Surabaya, 12 September 2019

Penulis

DAFTAR ISI

LEMBAR PERNYATAAN.....	ii
LEMBAR PENGESAHAN	iii
KETERANGAN REVISI	iv
SURAT PERNYATAAN ANTI PLAGIASI	v
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xvii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan	5
1.5 Manfaat	5
1.6 Sistematika Penulisan	6
BAB II	8
2.1 Penelitian Terdahulu	8
2.2 Dasar Teori.....	11
2.2.1 Twitter	11
2.2.2 Analisis Sentimen.....	11
2.2.3 <i>Text Mining</i>	12

2.2.4	Term frequency – inverse document frequency (TF-IDF)	12
2.2.5	Synthetic Minority Oversampling Technique (SMOTE)	14
2.2.6	Naïve Bayes	15
2.2.7	Evaluasi Performa Model Klasifikasi	17
2.2.8	Python	18
2.2.9	PHP	18
BAB III		19
3.1	Studi Literatur	20
3.2	Analisis Kebutuhan	20
3.2.1	Kebutuhan Data	21
3.2.2	Kebutuhan Hardware dan Software Perancangan	21
3.3	Membuat Model Klasifikasi	21
3.3.1	Pengumpulan Data	23
3.3.2	Penyaringan Data	23
3.3.3	Pelabelan Data	24
3.3.4	Preproses Data	24
3.3.5	Pembagian Data	31
3.3.6	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	31
3.3.7	Klasifikasi Data	31
3.3.8	Evaluasi Performa Model Klasifikasi	32
3.4	Analisis dan Perancangan Sistem	32
3.4.1	Analisis Sistem	33
3.4.2	Perancangan Antar Muka	33
3.5	Implementasi Sistem	33

BAB IV.....	34
4.1 Kebutuhan Sistem	34
4.1.1 Perangkat Keras (Hardware)	34
4.1.2 Perangkat Lunak (Software).....	35
4.2 Pembangunan Model.....	35
4.2.1 Pengumpulan Data	35
4.2.2 Penyaringan Data.....	36
4.2.3 Pelabelan Data	38
4.2.4 Pembagian Data.....	74
4.2.5 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	75
4.2.6 Klasifikasi Data	76
4.2.7 Evaluasi Performa Model Klasifikasi.....	87
4.3 Implementasi Sistem	93
BAB V	98
5.1 Kesimpulan	98
5.2 Saran.....	98
DAFTAR PUSTAKA.....	99

DAFTAR GAMBAR

Gambar 1.1	Perilaku Pengguna Internet.....	1
Gambar 1.2	Media Sosial yang Aktif di Indonesia.....	2
Gambar 3.1	Diagram Alir Penelitian.....	19
Gambar 3.2	Diagram Alir Penelitian (Lanjutan).....	20
Gambar 3.3	Diagram Alir Pembuatan Model.....	22
Gambar 3.4	Diagram Alir Tahapan Preproses.....	25
Gambar 4.1	<i>Source Code</i> Pengumpulan Data.....	35
Gambar 4.2	Contoh Hasil Pengumpulan Data.....	36
Gambar 4.3	Grafik Jumlah <i>Tweet</i>	36
Gambar 4.4	<i>Source Code Case Folding</i>	39
Gambar 4.5	<i>Source Code</i> Menghilangkan URL Situs Lain dan URL Gambar & Video.....	42
Gambar 4.6	<i>Source Code</i> Menghilangkan Nama Akun Atau <i>Mention</i>	45
Gambar 4.7	<i>Source Code</i> Menghilangkan Tagar atau <i>Hashtag</i>	47
Gambar 4.8	<i>Source Code</i> Menghilangkan Angka.....	50
Gambar 4.9	<i>Source Code</i> Menghilangkan Tanda Baca.....	52
Gambar 4.10	<i>Source Code</i> Menghilangkan Huruf Berulang.....	54
Gambar 4.11	<i>Source Code</i> Menghilangkan Huruf Tunggal.....	57
Gambar 4.12	<i>Source Code</i> Menghilangkan Spasi Berlebih.....	59
Gambar 4.13	<i>Source Code</i> Menghilangkan Baris Baru.....	61
Gambar 4.14	<i>Source Code</i> Mengubah Kata Singkatan.....	64
Gambar 4.15	<i>Source Code</i> Tokenisasi.....	66

Gambar 4.16	<i>Source Code Stopword Removal</i>	69
Gambar 4.17	<i>Source Code Stemming</i>	72
Gambar 4.18	<i>Source Code Pembobotan</i>	73
Gambar 4.19	Rincian Data Matriks Pembobotan.....	74
Gambar 4.20	Contoh Hasil Pembobotan Data.....	78
Gambar 4.21	<i>Source Code</i> Pembagian Data menggunakan metode <i>Hold Out</i> ...	75
Gambar 4.22	<i>Source Code</i> Pembagian Data menggunakan metode <i>Cross Validation</i>	75
Gambar 4.23	<i>Source Code</i> SMOTE.....	75
Gambar 4.24	<i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes dengan Metode Pembagian Data <i>Hold Out</i>	76
Gambar 4.25	<i>Source Code</i> Klasifikasi Data Bernoulli Naïve Bayes dengan Metode Pembagian Data <i>Hold Out</i>	77
Gambar 4.26	<i>Source Code</i> Klasifikasi Data Gaussian Naïve Bayes dengan Metode Pembagian Data <i>Hold Out</i>	77
Gambar 4.27	<i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Hold Out</i>	78
Gambar 4.28	<i>Source Code</i> Klasifikasi Data Bernoulli Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Hold Out</i>	78
Gambar 4.29	<i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Hold Out</i>	79

Gambar 4.30	<i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes dengan Metode Pembagian Data <i>Cross Validation</i>	80
Gambar 4.31	<i>Source Code</i> Klasifikasi Data Bernoulli Naïve Bayes dengan Metode Pembagian Data <i>Cross Validation</i>	80
Gambar 4.32	<i>Source Code</i> Klasifikasi Data Gaussian Naïve Bayes dengan Metode Pembagian Data <i>Cross Validation</i>	81
Gambar 4.33	<i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Cross Validation</i>	81
Gambar 4.34	<i>Source Code</i> Klasifikasi Data Bernoulli Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Cross Validation</i>	82
Gambar 4.35	<i>Source Code</i> Klasifikasi Data Gaussian Naïve Bayes menggunakan SMOTE dengan Metode Pembagian Data <i>Cross Validation</i>	82
Gambar 4.36	Probabilitas Kelas Negatif.....	84
Gambar 4.37	Probabilitas Kelas Positif.....	84
Gambar 4.38	Probabilitas Kelas Netral.....	84
Gambar 4.39	Banyak Kata Kelas Negatif.....	84
Gambar 4.40	Banyak Kata Kelas Positif.....	85
Gambar 4.41	Banyak Kata Kelas Netral.....	85
Gambar 4.42	Probabilitas Kata Kelas Negatif.....	85
Gambar 4.43	Probabilitas Kata Kelas Positif.....	86
Gambar 4.44	Probabilitas Kata Kelas Netral.....	86

Gambar 4.45	Probabilitas dengan Inplace Kelas Negatif.....	86
Gambar 4.46	Probabilitas dengan Inplace Kelas Positif.....	86
Gambar 4.47	Probabilitas dengan Inplace Kelas Netral.....	86
Gambar 4.48	Perkalian Kelas Positif.....	87
Gambar 4.49	Perkalian Kelas Negatif.....	87
Gambar 4.50	Perkalian Kelas Netral.....	87
Gambar 4.51	Diagram Hasil Akurasi Model Klasifikasi.....	89
Gambar 4.52	<i>Confusion Matrix</i> Skenario 11.....	90
Gambar 4.53	<i>Confusion Matrix</i> Skenario 11 (2).....	90
Gambar 4.54	<i>Confusion Matrix</i> Skenario 11 (3)	91
Gambar 4.55	<i>Confusion Matrix</i> Skenario 11 (4)	91
Gambar 4.56	<i>Confusion Matrix</i> Skenario 11 (5)	92
Gambar 4.57	<i>Confusion Matrix</i> Skenario 11 (6)	92
Gambar 4.58	<i>Confusion Matrix</i> Skenario 11 (7)	93
Gambar 4.59	Web Visualisasi.....	94
Gambar 4.60	Web Visualisasi (2).....	94
Gambar 4.61	Web Visualisasi (3).....	95
Gambar 4.62	Web Visualisasi (4).....	95
Gambar 4.63	Web Visualisasi (5).....	96
Gambar 4.64	Web Visualisasi (6).....	96
Gambar 4.65	Web Visualisasi (7).....	97

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu.....	8
Tabel 4.1	Contoh Hasil Penyaringan Data.....	37
Tabel 4.2	Contoh Hasil Pelabelan Data.....	38
Tabel 4.3	Contoh Hasil <i>Case Folding</i>	40
Tabel 4.4	Contoh Hasil Menghilangkan URL Situs Lain dan URL Gambar & Video.....	43
Tabel 4.5	Contoh Hasil Menghilangkan Nama Akun atau <i>Mention</i>	45
Tabel 4.6	Contoh Hasil Menghilangkan Tagar atau <i>Hashtag</i>	48
Tabel 4.7	Contoh Hasil Menghilangkan Angka.....	50
Tabel 4.8	Contoh Hasil Menghilangkan Tanda Baca.....	52
Tabel 4.9	Contoh Hasil Menghilangkan Huruf Berulang.....	55
Tabel 4.10	Contoh Hasil Menghilangkan Huruf Tunggal.....	57
Tabel 4.11	Contoh Hasil Menghilangkan Spasi Berlebih.....	59
Tabel 4.12	Contoh Hasil Menghilangkan Baris Baru.....	62
Tabel 4.13	Contoh Hasil Mengubah Kata Singkatan.....	64
Tabel 4.14	Contoh Hasil Tokenisasi.....	67
Tabel 4.15	Contoh Hasil <i>Stopword Removal</i>	69
Tabel 4.16	Contoh Hasil <i>Stemming</i>	72
Tabel 4.17	Contoh Hasil Klasifikasi Data.....	83
Tabel 4.18	Hasil Skenario	87

Judul Skripsi : ANALISIS SENTIMEN PEMINDAHAN IBU KOTA INDONESIA PADA TWITTER DENGAN MENGGUNAKAN METODE NAÏVE BAYES
Penulis : Erica Mas'udah
Pembimbing I : Eka Dyar Wahyuni, S.Kom, M.Kom
Pembimbing II : Amalia Anjani Arifiyanti, S.Kom, M.Kom

ABSTRAK

Indonesia merupakan pengguna sosial media dengan jumlah yang sangat besar. Sehingga, aktivitas-aktivitas bersosial media sangat banyak dilakukan oleh masyarakat. Pemindahan Ibu Kota Indonesia kini menjadi topik yang banyak dibicarakan dan masyarakat sering kali beropini dan dipublikasikan di media sosial seperti Twitter. Dari banyaknya opini, maka dapat menimbulkan suatu permasalahan, seperti perbedaan pendapat. Sehingga dari permasalahan tersebut dapat dimanfaatkan untuk dilakukan analisis sentimen mengenai pemindahan Ibu Kota Indonesia pada Twitter.

Sentimen yang dianalisis yakni berupa sentimen positif, negatif, dan netral. Untuk mendapatkan data *tweet* pada Twitter, perlu dilakukan pengumpulan data terlebih dahulu. Kemudian untuk melakukan analisis sentimen, data perlu diolah dengan *text mining*. Pada studi kasus ini, *text mining* yang dilakukan menggunakan metode Naïve Bayes. Metode Naïve Bayes yang digunakan ada 3 macam, dengan tujuan untuk mendapatkan nilai akurasi yang terbaik. Ditambah lagi dengan melakukan percobaan beberapa skenario.

Dari penelitian yang dilakukan, hasil akurasi terbaik didapatkan oleh skenario 11 yakni sebesar 68,10% yang mana menggunakan klasifikasi Bernoulli Naïve Bayes dengan pembagian data menggunakan metode *Cross Validation* dan melakukan tahap penyeimbangan data menggunakan metode SMOTE. Kemudian hasil dari analisis divisualisasikan di web agar mudah dipahami oleh orang awam.

Kata Kunci: *Data Mining, Ibu Kota Indonesia, Klasifikasi, Naïve Bayes, Sentimen Analisis, Twitter*