

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA SENTIMENT ANALYSIS OPINI
PEMBANGUNAN INFRASTRUKTUR DI MEDIA SOSIAL TWITTER**

SKRIPSI



Disusun Oleh :

Ignasius Frans De Sales Tyas Neno

17.21.1064

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA SENTIMENT ANALYSIS OPINI
PEMBANGUNAN INFRASTRUKTUR DI MEDIA SOSIAL TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana pada
Program Studi Informatika



Disusun oleh

Ignasius Frans De Sales Tyas Neno

17.21.1064

**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA SENTIMENT ANALYSIS OPINI
PEMBANGUNAN INFRASTRUKTUR DI MEDIA SOSIAL TWITTER**

yang dipersiapkan dan disusun oleh

Ignasius Frans De Sales Tyas Neno

17.21.1064

telah disetujui oleh Dosen Pembimbing Skripsi pada
tanggal 10 Juli 2019

Dosen Pembimbing,



Dina Maulina, M.Kom

NIK. 190302250

PENGESAHAN

SKRIPSI

**PERBANDINGAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN
SUPPORT VECTOR MACHINE PADA SENTIMENT ANALYSIS OPINI
PEMBANGUNAN INFRASTRUKTUR DI MEDIA SOSIAL TWITTER**

yang dipersiapkan dan disusun oleh
Ignasius Frans De Sales Tyas Neno

17.21.1064

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Juli 2019

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dina Maulina, M.Kom
NIK. 190302250

Mardhiya Havati, ST, M.Kom
NIK. 190302108

Bety Wulan Sari, M.Kom
NIK. 190302254



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 1 Agustus 2019

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si, M.T.

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 10 Agustus 2019



Ignasius Frans D.S.T.N

NIM 17.21.1064

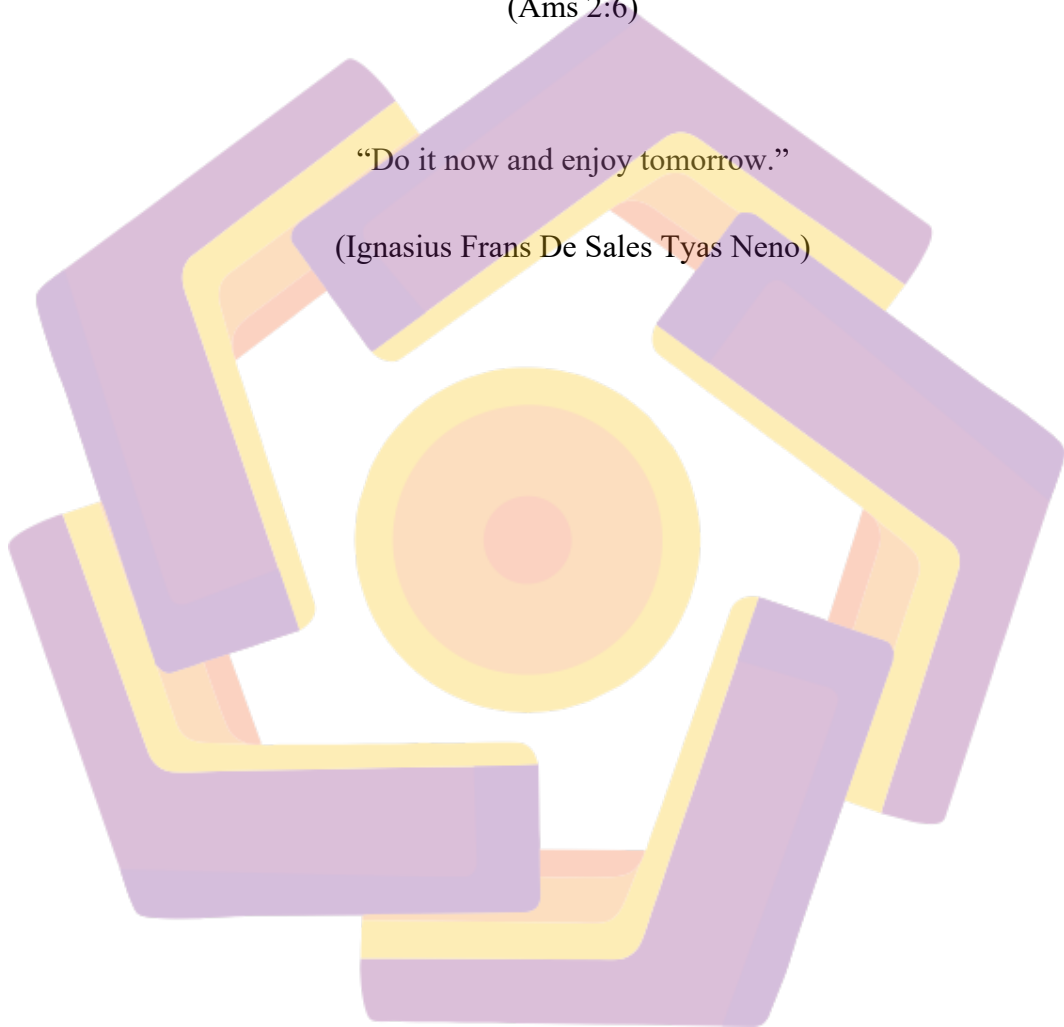
MOTTO

“Because the Lord gives wisdom, knowledge and intelligence come from His mouth. For, the Lord gives wisdom; knowledge and understanding come from His mouth.”

(Ams 2:6)

“Do it now and enjoy tomorrow.”

(Ignasius Frans De Sales Tyas Neno)



PERSEMBAHAN

Puji dan syukur penulis panjatkan kehadirat Tuhan Yesus Kristus yang telah melimpahkan rahmat serta berkatnya-Nya, sehingga penulis dapat menyelesaikan Skripsi yang berjudul **“Perbandingan Algoritma Naïve Bayes Classifier Dan Support Vector Machine Pada Sentiment Analysis Opini Pembangunan Infrastruktur Di Media Sosial Twitter”** dengan baik dan lancar.

Laporan Skripsi ini penulis ajukan sebagai salah satu syarat kelulusan program studi Strata satu Teknik Informatika Universitas Amikom Yogyakarta.

Pada kesempatan ini penulis ingin menyampaikan rasa hormat pad terimakasih kepada Bapak M. Suyanto, Prof. M.M. sebagai Rektor Universitas Amikom Yogyakarta dan Bapak Sudarmawan, MT selaku Ketua Jurusan S1 Teknik Informatika Universitas Amikom Yogyakarta serta Ibu Dina Maulina, M.Kom selaku dosen pembimbing.

Penulis menyadari bahwa implementasi dan laporan Skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik serta saran yang membangun sangat penulis harapkan demi kemajuan yang lebih baik dimasa mendatang. Semoga penelitian ini dapat memberikan tambahan ilmu pengetahuan dan manfaat pada semua pihak.

KATA PENGANTAR

Segala puji dan syukur kepada Allah Bapa yang maha kuasa atas rahmat dan berkat yang telah di berikan, sehingga penulis dapat menyelesaikan skripsi ini. Puji dan syukur juga tidak lupa penulis haturkan kepada Tuhan Yesus Kristus yang telah menguatkan mental dan keyakinan penulis dalam menyelesaikan skripsi yang telah di buat. Skripsi ini penulis persembahkan dengan rasa terimakasih yang besar kepada :

1. Kedua orang tua penulis Almarhum Bapak Thomas Neno Obe dan Ibu Sri Harjanti, S.Pd. serta kakak penulis Maria Emiliana D.T.N, S.Pd. yang telah memberikan semangat, mendidik, memotivasi, dan mendoakan penulis.
2. Ibu Dina Maulina, M.Kom selaku dosen pembimbing yang sudah penulis anggap sebagai orang tua yang telah membimbing dan memberikan arahan kepada penulis semalam pengerjaan skripsi ini.
3. Sahabat-sahabat seperjuangan kelas transfer angkatan 2017 yang terkumpul dari berbagai daerah di Indonesia, khususnya Try, Aurora, Lia, Umar, Nisya, Rahma, Iin, Intan, dan Cholif. Sahabat-sahabat sepergabatan yang tiba-tiba mengajak jalan kemana-mana.
4. Teman seperantauan Gea yang telah memberikan masukan pada skripsi penulis dan masih banyak lagi teman-teman yang tidak bisa di sebutkan satu persatu oleh penulis.

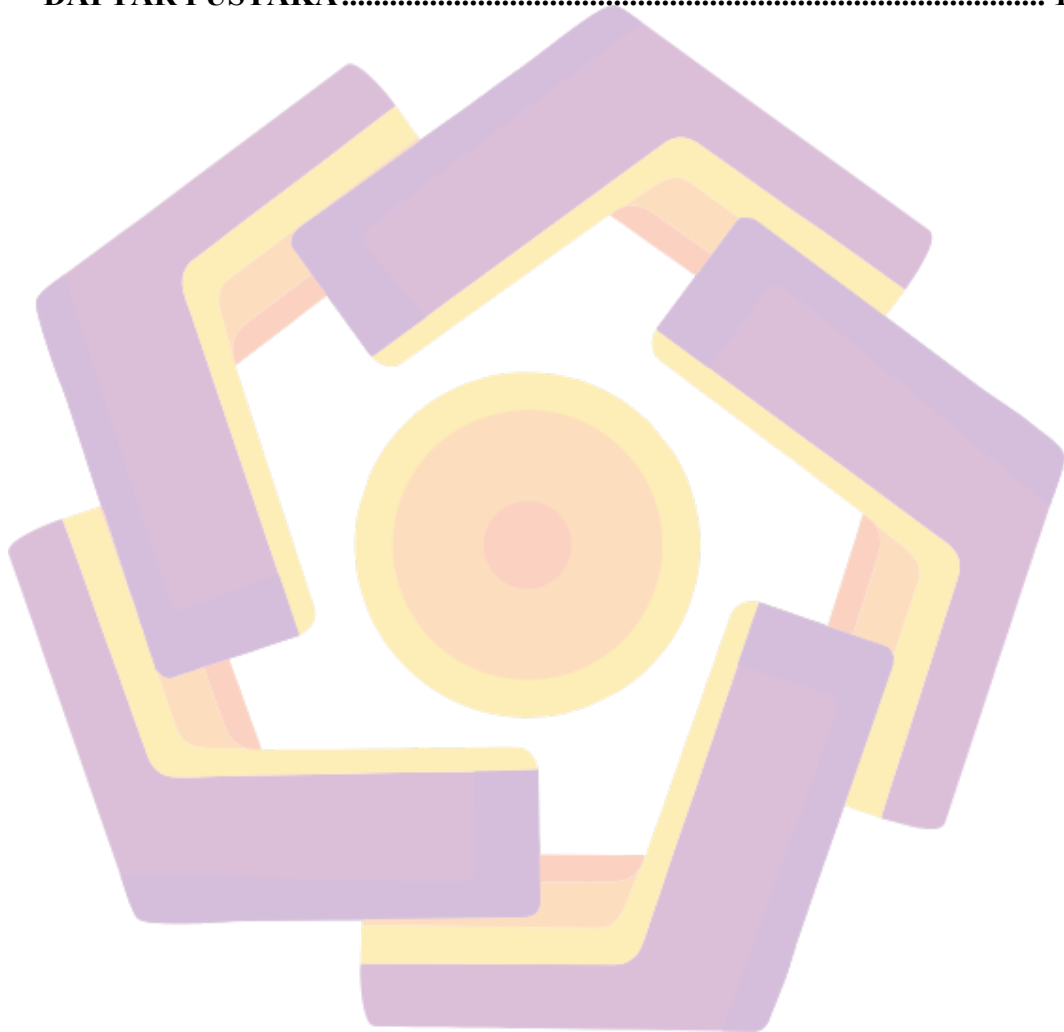
DAFTAR ISI

PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN	v
MOTTO	iv
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTARR TABEL	xiv
DAFTAR GAMBAR	xiii
INTISARI	xiv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	4
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	5
1.5 Manfaat Penelitian.....	6
1.6 Metodologi Penelitian	7
1.6.1 Metode Pengumpulan Data.....	7
1.6.2 Metode Analisis	8
1.6.3 Perancangan	8
1.6.4 Implementasi.....	10
1.6.5 Testing.....	10
1.7 Sistematika Penulisan.....	10
BAB II LANDASAN TEORI	12
2.1 Tinjauan Pustaka	12
2.2 Data Mining.....	13
2.2.1 Analisis Sentimen	14
2.2.2 Klasifikasi	14
2.2.3 Algoritma	15

2.2.3.1	Naïve Bayes Classifier.....	16
2.2.3.2	Support Vector Machine.....	17
2.2.4	Crawling Data	19
2.2.5	Pre-Processing.....	19
2.2.5.1	Tokenizing	21
2.2.5.2	Stopword Removal.....	22
2.2.5.3	Stemming.....	23
2.2.5.4	TF-IDF	23
2.2.5.5	Feature List	23
2.2.5.6	Confusion Matriks	24
2.2.6	Cross Validation.....	24
2.2.7	Bahasa Pemrograman R.....	24
2.2.7.1	R-Studio.....	25
2.2.8	Rapid Miner.....	26
2.2.9	Flowchart	27
BAB III METODOLOGI PENELITIAN		29
3.1	Gambaran Umum Jalannya Proses Penelitian.....	29
3.2	Analisis Kebutuhan Sistem.....	31
3.2.1	Kebutuhan Fungsional	31
3.2.2	Kebutuhan Non-Fungsional	31
3.2.2.1	Analisis Kebutuhan Perangkat Keras (Hardware).....	31
3.2.2.2	Analisis Kebutuhan Perangkat Lunak (Software)	32
3.3	Analisa Penelitian.....	32
3.3.1	Pengumpulan Data	32
3.3.2	Pelabelan Manual Data Training.....	34
3.3.3	Tokenizing	37
3.3.4	Stopword Removal.....	40
3.3.5	Stemming	44
3.3.6	Feature List	47
3.3.6.1	Term Positif	48
3.3.6.2	Term Negatif.....	51

3.3.7	Data Testing	53
3.3.7.1	Pelabelan Data Testing	54
3.3.7.2	Tokenizing Data Testing.....	55
3.3.7.3	Stopword Removal Data Testing.....	56
3.3.7.4	Stemming Data Testing	57
3.3.8	TF-IDF	59
3.3.9	Analisis Naïve Bayes Classifier.....	60
3.3.9.1	Nilai Prior	64
3.3.9.2	Nilai Posterior.....	65
3.3.9.3	Hasil Klasifikasi.....	69
3.3.9.4	Confusion Matriks	70
3.3.10	Analisis Support Vector Machine	71
3.4	Analisis Dataset.....	76
3.4.1	Confusion Matriks Pada Naïve Bayes Classifier	104
3.4.1	Confusion Matriks Pada Support Vector Machine.....	104
BAB IV HASIL DAN IMPLEMENTASI		107
4.1	Implementasi Crawling Data.....	107
4.1.1	Library R Programming	107
4.1.2	Pemanggilan Library	107
4.1.3	Pemasangan API	108
4.1.4	Crawling Data	109
4.1.5	Cleaning Data.....	109
4.1.6	Export Excel.....	111
4.2	Implementasi	111
4.2.1	Pelabelan Manual	111
4.2.2	Skema Alur Proses Klasifikasi.....	112
4.2.3	Import Dataset.....	114
4.2.4	Set Role Model.....	116
4.2.5	Data Training Dan Data Testing	117
4.2.6	Select Attribute.....	118
4.2.7	Preprocessing	118

4.2.8	Cross Validation.....	119
4.3	Hasil Klasifikasi Dan Perbandingan Hasil	120
BAB V PENUTUP		122
5.1	Kesimpulan.....	122
5.2	Saran.....	122
DAFTAR PUSTAKA		124



DAFTAR TABEL

Tabel 2.1 Tabel Flowchart.....	27
Tabel 3.1 Tabel Crawling Data.....	32
Tabel 3.2 Tabel Pelabelan	34
Tabel 3.3 Tabel <i>Tokenizing</i>	37
Tabel 3.4 Tabel Stopword	41
Tabel 3.5 Tabel Stemming.....	45
Tabel 3.6 Tabel Hasil Stemming.....	47
Tabel 3.7 Tabel Term Positif.....	48
Tabel 3.8 Tabel Term Negatif	51
Tabel 3.9 Tabel Pelabelan Data Testing.....	54
Tabel 3.10 Tabel Tokenizing Data Testing	55
Tabel 3.11 Tabel Stopword Data Testing.....	56
Tabel 3.12 Tabel Stemming Data Testing.....	58
Tabel 3.13 Tabel Pengelompokan Data Training	60
Tabel 3.14 Tabel Perhitungan Posterior Tweet 1.....	66
Tabel 3.15 Tabel Perhitungan Posterior Tweet 2.....	67
Tabel 3.16 Tabel Perhitungan Posterior Tweet 3.....	67
Tabel 3.17 Tabel Perhitungan Posterior Tweet 4.....	68
Tabel 3.18 Tabel Hasil Klasifikasi	69
Tabel 3.19 Perhitungan Confusion Matriks	70
Tabel 3.20 Tabel Contoh Data Vector.....	71
Tabel 3.21 Tabel Dataset	77
Tabel 4.1 Tabel Fungsi Operator.....	113
Tabel 4.2 Tabel Fungsi Operator Preprocessing.....	119
Tabel 4.3 Perbandingan Nilai Akurasi.....	121

DAFTAR GAMBAR

Gambar 2.1 Tampilan R-Studio	25
Gambar 2.2 Tampilan Rapid Miner	26
Gambar 3.1 Alur Proses Penelitian	29
Gambar 3.2 Tampilan Kordinat.....	72
Gambar 3.3 Support Vector.....	73
Gambar 4.1 Instalasi Packages R	107
Gambar 4.2 Library R.....	108
Gambar 4.3 Setup API.....	108
Gambar 4.4 Crawling Data	109
Gambar 4.5 Cleaning Tweet	110
Gambar 4.6 Export Excel.....	111
Gambar 4.7 Dataset Mentah	112
Gambar 4.8 Skema Klasifikasi	113
Gambar 4.9 Step 1 Import data	114
Gambar 4.10 Step 2 Import data	115
Gambar 4.11 Step 3 Import data	115
Gambar 4.12 Parameter Operator Set Role Model	116
Gambar 4.13 Parameter Operator Data Training dan Data Testing	117
Gambar 4.14 Parameter Operator Select Attribute	118
Gambar 4.16 Operator Preprocessing	118
Gambar 4.17 Validasi SVM	119
Gambar 4.18 Validasi NBC.....	120
Gambar 4.18 Akurasi NBC	120
Gambar 4.20 Validasi SVM	121

INTISARI

Tujuan dari skripsi ini adalah menganalisa perbandingan model pembelajaran dari dua algoritma yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC) yang di implementasikan kepada objek sentimen masyarakat Indonesia di media sosial *Twitter*. Selama ini telah di ketahui bahwa kebanyakan sentimen analisis yang telah di lakukan hanya menggunakan model pembelajaran tertentu. Diperlukan perbandingan terhadap model lain guna menemukan nilai akurasi terbaik dari setiap model yang di pilih.

Pada saat ini media sosial sebagai tempat untuk beropini bisa di pandang juga sebagai tempat penyimpanan data yang memiliki pola tertentu. Sentimen analisis akan menggunakan model tertentu untuk menghasilkan informasi yang lebih terproses agar data yang di sajikan bisa mensimpulkan pola yang berkembang di masyarakat.

Penelitian ini akan mensimpulkan perbandingan hasil akurasi dari NBC dan SVM serta melihat pola opini yang berkembang di masyarakat terhadap pembangunan infrastruktur yang di jalankan oleh pemerintah Indonesia.

Kata Kunci : Sentimen Analisis, Naïve Bayes Classifier, Support Vector Machine, Pembangunan Infrastruktur, Klasifikasi.

ABSTRACT

The purpose of this thesis is to analyze the comparison of learning models from two algorithms namely Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC) which are implemented to the object of Indonesian public sentiment on Twitter social media. So far it has been known that most of the analytical sentiments that have been carried out only use certain learning models. Comparison of other models is needed in order to find the best accuracy value for each chosen model.

At this time social media as a place for opinion can also be seen as a data storage area that has a certain pattern. Sentiment analysis will use certain models to produce more processed information so that the data presented can infer patterns that develop in the community.

This study will conclude a comparison of the results of accuracy from NBC and SVM and see the pattern of opinion that develops in the community towards infrastructure development carried out by the Indonesian government.

Keyword : *Sentiment Analysis, Naïve Bayes Classifier, Support Vector Machine, Infrastructure Development, Classification.*