

**OPTIMALISASI SENTIMEN ANALISIS TENTANG BALAPAN
FORMULA E DI JAKARTA MENGGUNAKAN NAÏVE BAYES DAN
PARTICLE SWARM OPTIMIZATION**

SKRIPSI



disusun oleh

Ireneu Hanjan Da Silva

17.11.1522

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

2023

**OPTIMALISASI SENTIMEN ANALISIS TENTANG BALAPAN
FORMULA E DI JAKARTA MENGGUNAKAN NAÏVE BAYES DAN
PARTICLE SWARM OPTIMIZATION**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Ireneu Hanjan Da Silva

17.11.1522

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMALISASI SENTIMEN ANALISIS TENTANG BALAPAN
FORMULA E DI JAKARTA MENGGUNAKAN NAÏVE BAYES DAN
PARTICLE SWARM OPTIMIZATION**

yang dipersiapkan dan disusun oleh

Ireneu Hanjan Da Silva

17.11.1522

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 05 Desember 2022

Dosen Pembimbing,

Ferlan Fauzi Abdulloh, M.Kom

NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

**OPTIMALISASI SENTIMEN ANALISIS TENTANG BALAPAN
FORMULA E DI JAKARTA MENGGUNAKAN NAÏVE BAYES DAN
PARTICLE SWARM OPTIMIZATION**

yang disusun dan diajukan oleh

Ireneu Hanjan Da Silva
17.11.1522

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Desember 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Irma Rofni Wulandari, S.Pd., M.Eng Penguji 1
NIK. 190302329

Sharazita Dyah Anggita, M.Kom Penguji 2
NIK. 190302285

Ferian Fauzi Abdullah, M.Kom Penguji 3
NIK. 190302276

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

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN 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 tanggung jawab saya pribadi.

Yogyakarta, Desember 2022



Ireneu Hanjan Da Silva
NIM. 17.11.1522

HALAMAN MOTTO

“Segala perkara dapat kutanggung di dalam Dia yang memberi kekuatan kepadaku”

~ Filipi 4:13 ~

“Banyaklah rancangan di hati manusia, tetapi keputusan Tuhanlah yang terlaksana”

~ Amsal 19:21 ~



HALAMAN PERSEMBAHAN

Puji dan Syukur atas rahmat dan berkat yang telah diberikan oleh Tuhan Yang Maha Kuasa sehingga karya ini dapat terselesaikan dengan baik dan lancar.

Skripsi ini dipersembahkan kepada :

1. Tuhan yang Maha Kuasa yang memberikan segenap kekuatan dan kemampuan untuk dapat menyelesaikan skripsi ini dan memudahkan jalan agar lancar dalam mengerjakan dan dalam proses bimbingan.
2. Kedua Orang Tua saya, Ibu Fransisca Hanjan De Jesus dan Bapak Mario Eduardo Pereira Da Silva yang tiada hentinya mendoakan, memberi dukungan dan semangat, serta berjuang tak kenal lelah untuk membantu dalam menyelesaikan pendidikan.
3. Kakaku Ledy, Alou, Asalu, Abeno Claudia yang telah memberikan semangat dan dukungan.
4. Arsita Sary Tefa yang selalu memberikan semangat dan dukungan.
5. Keluarga besar kelas Informatika 09 2017 yang sangat luar biasa.

KATA PENGANTAR

Puji dan syukur panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, berkat dan kekuatan sehingga peneliti dapat menyelesaikan skripsi yang berjudul optimalisasi sentimen analisis tentang balapan formula e di jakarta menggunakan naïve bayes dan particle swarm optimization.

Skripsi ini dibuat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini peneliti mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya sampai sejauh ini.
2. Bapak Prof.Dr.M.Suyanto,MM. selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta,S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Ferian Fauzi Abdulloh, M.Kom. selaku dosen pembimbing yang selalu bijaksana dalam memberikan bimbingan, nasehat serta waktunya selama pembuatan skripsi ini.
5. Dosen penguji dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
6. Kedua orang tua dan keluarga yang telah mendoakan, mendukung dan memberikan semangat.
7. Orang-orang sekitar yang senantiasa mendoakan, memberi semangat dan bantuan.
8. Teman-teman IF 09 2017 yang telah menemani selama proses perkuliahan.
9. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Tuhan Yang Maha Kuasa memberikan balasan yang lebih kepada semua yang telah ikut membantu peneliti hingga menyelesaikan skripsi ini.Demi perbaikan selanjutnya,saran dan kritik membangun diterima dengan senang hati dan terima kasih.Semoga skripsi ini dapat bermanfaat bagi peneliti dan kita semua.

Yogyakarta, Desember 2022

Ireneu Hanjan Da Silva

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PERNYATAAN	iii
HALAMAN PENGESAHAN	iii
HALAMAN MOTTO	iv
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Twitter	14
2.3 Analisis Sentimen	14
2.4 Preprocessing	15
2.5 Term Frequency – Inverse Document Frequency	16
2.6 Naïve Bayes Classifier	18
2.7 Particle Swarm Optimization	19
2.8 K-fold Cross Validation	19
2.9 Confussion Matrix	20

2.10 Python	21
2.11 Jupyter Notebook	22
BAB III METODE PENELITIAN	23
3.1 Objek Penelitian	23
3.2 Alur Penelitian	23
3.3 Bahan dan Metode	24
3.3.1 Crawling Data	24
3.3.2 Software	24
3.3.3 Preprocessing	25
3.3.4 Term Frequency – Inverse Document Frequency	25
3.3.5 Naïve Bayes	26
3.3.5 Particle Swarm Optimization	28
3.4 Evaluasi	29
3.5 Skenario Hasil Pengujian	30
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Hal-hal Pengambilan Data	31
4.2 Preprocessing	33
4.2.1 Visualisasi Wordcloud	36
4.2.2 Grafik Sentimen	37
4.3 Pembobotan TF-IDF	37
4.4 Evaluasi Hasil Pengujian	38
4.5 Hasil Peningkatan Accuray, Precision dan Recall	42
4.5 Hasil Perbandingan	43
BAB V PENUTUP	44
5.1 Kesimpulan	44
5.2 Saran	44
REFERENSI	45

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Confusion Matrix	20
Tabel 3. 1 Perhitungan TF-IDF	26
Tabel 3. 2 Perhitungan Kategori	26
Tabel 3. 3 Perhitungan kemunculan term kata.....	28
Tabel 3. 4 Perhitungan Confusion Matrix.....	29
Tabel 4. 1 Hasil Naïve Bayes tanpa PSO dengan K fold 10.....	38
Tabel 4. 2 Hasil NB-PSO dengan K fold 10.....	39
Tabel 4. 3 Hasil Naïve Bayes tanpa PSO dengan K fold 5.....	40
Tabel 4. 4 Hasil NB-PSO dengan K fold 5.....	40
Tabel 4. 5 Hasil Naïve Bayes tanpa PSO dengan K fold 3.....	41
Tabel 4. 6 Hasil NB-PSO dengan K fold 3.....	42
Tabel 4. 7 Tabel Perhitungan Perbandingan.....	42

DAFTAR GAMBAR

Gambar 2. 1 Alur Preprocessing	15
Gambar 2. 2 Alur Pembobotan TF-IDF	17
Gambar 3. 1 Alur Penelitian	24
Gambar 4. 1 crawling menggunakan sntwitter	31
Gambar 4. 2 Tampilan Jupyter Notebook	32
Gambar 4. 3 Hasil data crawling dan labelling	32
Gambar 4. 4 kode Program Case folding	33
Gambar 4. 5 kode program proses cleaning	34
Gambar 4. 6 Tampilan code Tokenization	34
Gambar 4. 7 Tampilan code stemming	35
Gambar 4. 8 Kode Program Proses Stop Removal	35
Gambar 4. 9 Tampilan Preprocessing	36
Gambar 4. 10 Tampilan WordCloud	36
Gambar 4. 11 Tampilan Grafik Sentimen	37
Gambar 4. 12 Hasil Pembobotan TF-IDF	38
Gambar 4. 13 Hasil Naïve Bayes k fold 10	39
Gambar 4. 14 Hasil NB-PSO k fold 10	39
Gambar 4. 15 Hasil Naïve Bayes k fold 5	40
Gambar 4. 16 Hasil NB-PSO k fold 5	41
Gambar 4. 17 Hasil Naïve Bayes k fold 3	41
Gambar 4. 18 Hasil NB-PSO k fold 3	42
Gambar 4. 19 Grafik Hasil Perbandingan	43

INTISARI

Analisis sentimen merupakan sebuah teknik untuk proses mengidentifikasi sentimen atau pendapat dalam berbagai sumber data dari internet dan beragam platform media sosial dan kemudian mengkategorikannya kedalam sentimen positif, negatif dan netral. Penelitian ini akan dilakukan optimalisasi sentimen analisis tentang balapan formula e di jakarta yang diutarakan melalui media sosial twitter.

Penelitian akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen tweet dimulai dari tahap crawling data dari media sosial twitter dengan memanfaatkan *snwtwitter* atau modul yang di sediakan twitter untuk dapat menarik data dari twitter dengan keywords Formula E Jakarta, preprocessing, pembobotan kata dengan TF-IDF sampai tahap analisis sentimen dengan menggunakan algoritma Naive Bayes dan Particle Swarm Optimization dengan menggunakan software jupyter notebook. Serta bagaimana mengukur kualitas hasil analisis sentimen dari Nilai K fold yang berbeda. Particle Swarm Optimization merupakan teknik optimasi yang digunakan untuk menerapkan dan memodifikasi beberapa parameter dan meningkatkan bobot atribut.

Dalam Penelitian ini, hasil perhitungan metode Naive Bayes dan Particle Swarm Optimization mendapatkan nilai hasil tertinggi dengan akurasi 95,77%, precision 95,85%, recall 99,87% dengan nilai k fold 10.

Kata kunci: analisis sentiment, TF-IDF, Naive Bayes, PSO.

ABSTRACT

Sentiment analysis is a technique to identify sentiments or opinions in various data sources from the internet and various social media platforms and then categorize them into positive, negative and neutral sentiments. This research will optimize sentiment analysis about the formula e race in Jakarta expressed through twitter social media.

This research will discuss the stages that are passed to perform the tweet sentiment analysis process starting from the data crawling stage from twitter social media by utilizing snwtwitter or the module provided by twitter to be able to pull data from twitter with the keyword Formula E Jakarta, preprocessing, word weighting with TF-IDF to the sentiment analysis stage using the Naive bayes algorithm and Particle Swarm Optimization using jupyter notebook software. Particle Swarm Optimization is an optimization technique used to apply and modify several parameters and increase attribute weights.

In this study, the calculation results of the Naïve Bayes and Particle Swarm Optimization methods get the highest result value with 95.77% accuracy, 95.85% precision, 99.87% recall. with a k fold value of 10.

Keyword: sentiment analysis,TF-IDF,Nave Bayes,PSO.