

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
10 DESTINASI WISATA PRIORITAS DI INDONESIA
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI



disusun oleh

Wibi Chamim Mushodiq

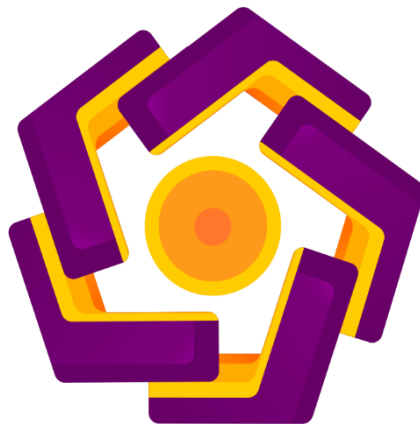
16.11.0068

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
10 DESTINASI WISATA PRIORITAS DI INDONESIA
MENGUNAKAN METODE NAIVE BAYES**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Wibi Chamim Mushodiq

16.11.0068

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP 10 DESTINASI WISATA PRIORITAS DI INDONESIA MENGUNAKAN METODE NAIVE BAYES

yang dipersiapkan dan disusun oleh

Wibi Chamim Mushodiq

16.11.0068

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 Maret 2020

Dosen Pembimbing,

Andi Sunyoto, M.Kom., Dr.

NIK. 190302052

PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
10 DESTINASI WISATA PRIORITAS DI INDONESIA
MENGUNAKAN METODE NAIVE BAYES**

yang dipersiapkan dan disusun oleh

Wibi Chamim Mushodiq

16.11.0068

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Maret 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Andi Sunyoto, M.Kom., Dr.
NIK. 190302052

Sumarni Adi, S.Kom, M.Cs
NIK. 190302256

Hartatik, S.T, M.G
NIK. 190302232

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Maret 2020

KETUA STMIK AMIKOM YOGYAKARTA

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

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, 18 Maret 2020

Meterai
Rp. 6.000

Wibi Chamim Mushodiq

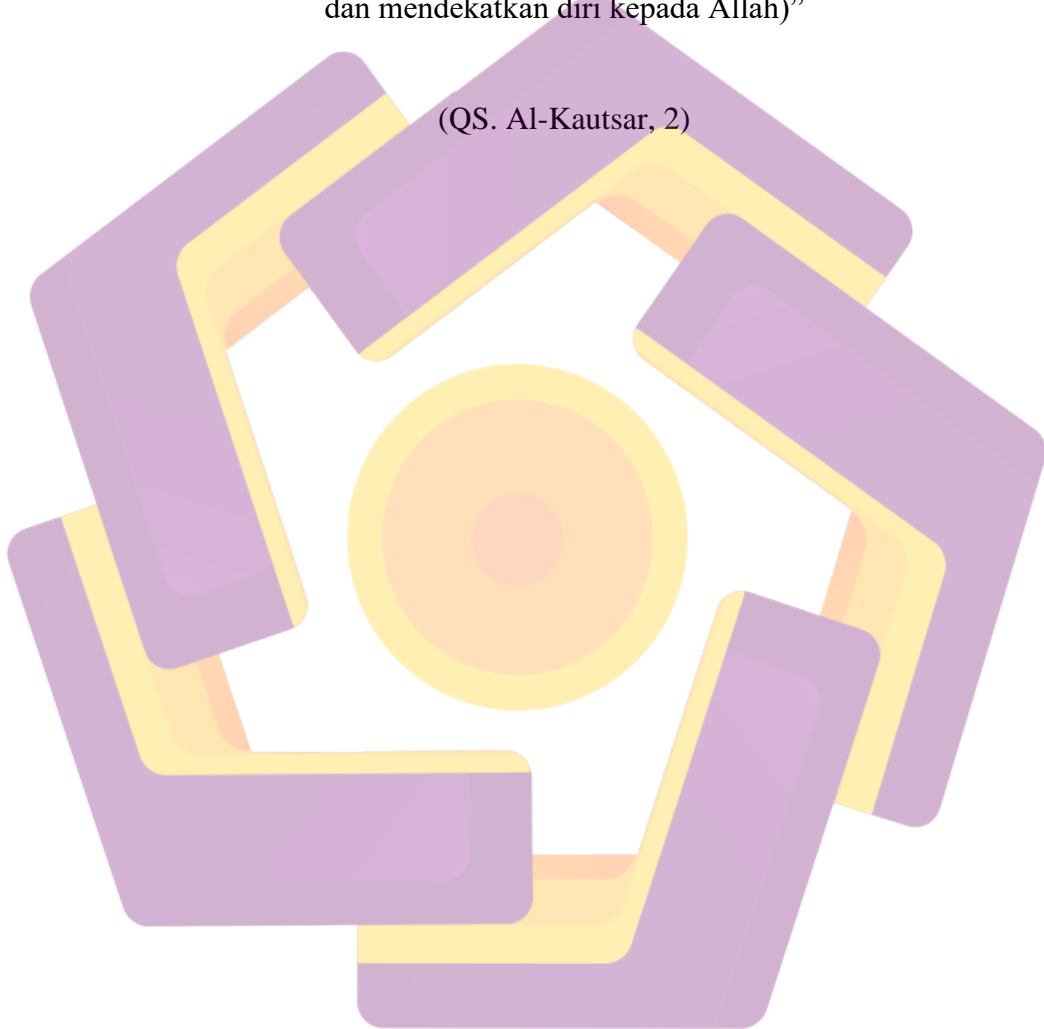
NIM. 16.11.0068

MOTTO

فَصَلِّ لِرَبِّكَ وَأَنْحَرْ

“Maka laksanakanlah shalat karena Tuhanmu, dan berkurbanlah (sebagai ibadah dan mendekatkan diri kepada Allah)”

(QS. Al-Kautsar, 2)



PERSEMBAHAN

Alhamdulillah, segala puji bagi Allah, Rabb semesta alam yang telah memberikan segala kenikmatan yang tidak bisa hitung jumlahnya. Shalawat serta salam semoga senantiasa ditujukan bagi Rasulullah ﷺ, keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Terima kasih kepada orang tua saya tercinta, Sumarno dan Sri Yuliyani yang telah mendidik, mendukung, menemani saya setiap saat, dan setiap hari mendo'akan saya. Kepada adik saya Fatchuz Zanuwar Rifa'i dan Ikhsanaw Wataufiqi Fitria terima kasih atas segala dukungannya semoga keluarga kita kelak dikumpulkan di surga – Nya.

Terima kasih kepada dosen – dosen Universitas Amikom Yogyakarta yang telah sabar dan ikhlas dalam mendidik saya hingga mampu melewati seluruh ujian mata kuliah, terutama kepada Bapak Andi Sunyoto, M.Kom., Dr. semoga ilmu yang beliau berikan dapat diterima dengan baik dan bermanfaat dikemudian hari.

Terima kasih untuk semua teman-teman kelas Informatika 02, teman-teman X-EA Entertainment, teman-teman dekat Febrilia Endah Kurniawati, Arif Budiarto, Arif Gunawan, Elvan Azikri Winanda, Yongki Budi Setyawanda, Muhammad Jauharul Ramadhan, sahabat-sahabatku Zakaria Mursid, Aditya Bayu Febriawan, Bagus Edi Prayogo, Fadhil Syahputra, Haris Eko Avrianto, Lukman Haryanto, Muhammad Arif Sadewa, dan Tasrifin yang telah menemani penulis berjuang di Yogyakarta. Semoga kelak di akhirat kita termasuk golongan orang yang mendapat naungan Allah dimana tidak ada naungan kecuali naungan – Nya.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Alhamdulillah *rabbil'Alamin*. Segala puji bagi Allah SWT yang telah memberikan rahmat, hidayah, dan ridha-Nya, sehingga penulis dapat menyelesaikan skripsi ini, dan studi penulis di jurusan Informatika Universitas Amikom Yogyakarta dengan baik dan lancar.

Dalam menyelesaikan skripsi ini, banyak pihak yang telah memberikan bantuan baik secara moril, nasihat, dan semangat maupun materil. Atas segala bantuan yang telah diberikan, penulis ingin menyampaikan doa dan ucapan terimakasih yang sedalam – dalamnya kepada :

1. M. Suyanto, Prof., Dr., M.M. selaku rektor Universitas Amikom Yogyakarta .
2. Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta
3. Sudarmawan, S.T., M.T. selaku ketua Jurusan Informatika Universitas Amikom Yogyakarta
4. Andika Agus Slameto, M.Kom, selaku dosen wali selama saya menempuh pendidikan S1 di Universitas Amikom Yogyakarta
5. Andi Sunyoto, M.Kom., Dr., selaku dosen pembimbing skripsi, yang telah banyak memberikan pengarahan

6. Segenap civitas akademika jurusan Informatika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.
7. Semua pihak yang ikut membantu dalam menyelesaikan skripsi ini baik berupa materiil maupun moril.

Peneliti menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan peneliti berharap semoga dapat diperbaiki dan disempurnakan oleh peneliti selanjutnya dan semoga skripsi ini dapat memberikan manfaat.

Aamiin ya rabbal 'alamin.

Wassalamualaikum Wr.Wb.

Yogyakarta, 18 Maret 2020

Wibi Chamim Mushodiq

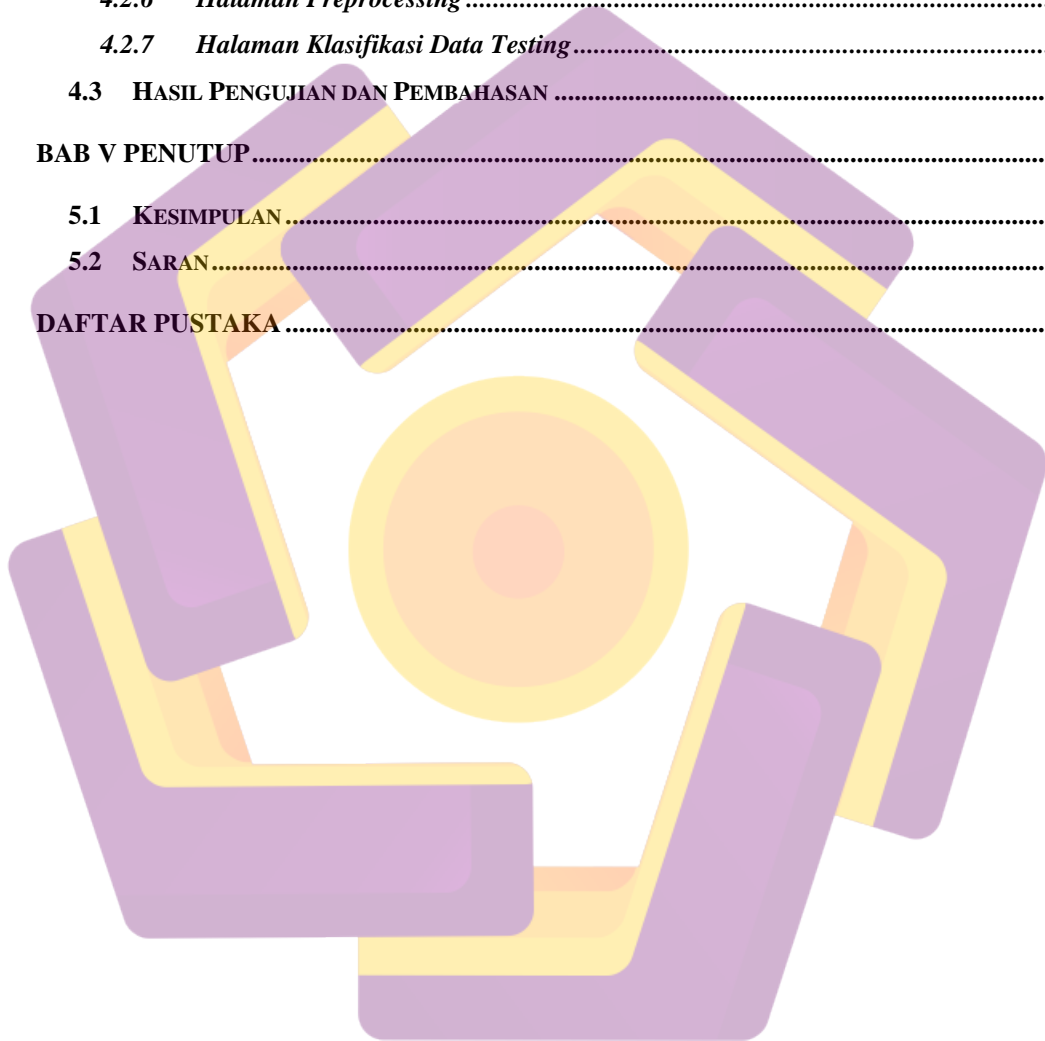
NIM. 16.11.0068

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN.....	II
PENGESAHAN	III
PERNYATAAN.....	III
MOTTO	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	XII
DAFTAR GAMBAR.....	XIII
DAFTAR ISTILAH	XV
<i>ABSTRACT</i>	XVIII
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	4
1.4 MAKSUD DAN TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN	5
1.7 SISTEMATIKA PENULISAN.....	6
BAB II LANDASAN TEORI	8
2.1 KAJIAN PUSTAKA.....	8
2.2 LANDASAN TEORI.....	9
2.2.1 <i>Twitter</i>	9
2.2.2 <i>Analisis Sentimen</i>	10
2.2.3 <i>Preprocessing</i>	11
2.2.4 <i>Pembobotan Kata (Term Weighting)</i>	13
2.2.5 <i>Naïve Bayes Classifier</i>	16
2.2.6 <i>Confusion Matrix</i>	17

BAB III METODE PENELITIAN	19
3.1 ALAT DAN BAHAN PENELITIAN	19
3.1.1 <i>Perangkat Lunak (Software)</i>	19
3.1.2 <i>Perangkat Keras (Hardware)</i>	19
3.2 ALUR PENELITIAN.....	20
3.2.1 <i>Metode Pengumpulan Dataset 10 Destinasi Wisata</i>	20
3.2.2 <i>Metode Pengumpulan Dataset</i>	21
3.2.3 <i>Preprocessing</i>	22
3.2.3.1 <i>Case Folding</i>	23
3.2.3.2 <i>Cleansing</i>	23
3.2.3.3 <i>Konversi Emoji</i>	24
3.2.3.4 <i>Konversi Negasi</i>	24
3.2.3.5 <i>Normalisasi</i>	25
3.2.3.6 <i>Stopword Removal</i>	25
3.2.3.7 <i>Stemming</i>	26
3.2.3.8 <i>Tokenisasi</i>	26
3.2.4 <i>Pembobotan Kata (TF-IDF)</i>	27
3.2.5 <i>Proses Training</i>	30
3.2.6 <i>Proses Testing</i>	33
3.2.7 <i>Perancangan Desain Antarmuka</i>	35
BAB IV HASIL DAN PEMBAHASAN.....	40
4.1 RANCANGAN SISTEM.....	40
4.1.1 <i>Pengumpulan Data (Crawling)</i>	40
4.1.2 <i>Pembuatan data latih</i>	41
4.1.3 <i>Praproses Data</i>	41
4.1.3.1 <i>Case Folding</i>	42
4.1.3.2 <i>Cleansing</i>	42
4.1.3.3 <i>Konversi Emoji</i>	43
4.1.3.4 <i>Konversi Negasi</i>	44
4.1.3.5 <i>Normalisasi</i>	45
4.1.3.6 <i>Tokenisasi</i>	47
4.1.3.7 <i>Stemming</i>	47
4.1.4 <i>Implementasi Pembobotan TF-IDF</i>	48
4.1.5 <i>Implementasi Klasifikasi</i>	49
4.1.5.1 <i>Implementasi Fungsi Pembobotan Naive Bayes Classifier</i>	49
4.1.5.2 <i>Implementasi Klasifikasi Data Testing</i>	50

4.2	IMPLEMENTASI ANTARMUKA	54
4.2.1	<i>Halaman Dashboard</i>	54
4.2.2	<i>Halaman Dataset Training</i>	54
4.2.3	<i>Halaman Bobot TF-IDF</i>	55
4.2.4	<i>Halaman Bobot Naive Bayes</i>	56
4.2.5	<i>Halaman Dataset Testing</i>	57
4.2.6	<i>Halaman Preprocessing</i>	57
4.2.7	<i>Halaman Klasifikasi Data Testing</i>	58
4.3	HASIL PENGUJIAN DAN PEMBAHASAN	59
BAB V	PENUTUP	62
5.1	KESIMPULAN	62
5.2	SARAN	63
	DAFTAR PUSTAKA	65



DAFTAR TABEL

Tabel 2.1	Confusion Matrix Positif, Negatif, dan Netral	16
Tabel 3.1	Daftar Kata Kunci	20
Tabel 3.2	Struktur Tabel Data Training	20
Tabel 3.3	Contoh Hasil Case Folding	22
Tabel 3.4	Contoh Hasil Cleansing	23
Tabel 3.5	Contoh Hasil Convert Emoji	23
Tabel 3.6	Contoh Hasil Convert Negation	24
Tabel 3.7	Contoh Hasil Normalisasi	24
Tabel 3.8	Contoh Hasil Stopword Removal	25
Tabel 3.9	Contoh Hasil Stemming	25
Tabel 3.10	Contoh Hasil Tokenisasi	26
Tabel 3.11	Contoh koleksi data	26
Tabel 3.12	Hasil Preprocessing koleksi data	27
Tabel 3.13	perhitungan bobot pada masing-masing dokumen	28
Tabel 3.14	Contoh data training	29
Tabel 3.15	Hasil perhitungan probabilitas	31
Tabel 3.16	Contoh data testing	32
Tabel 3.17	Contoh hasil perhitungan probabilitas	33
Tabel 4.1	Komposisi Sentimen data training	40
Tabel 4.2	Jumlah data pengujian akurasi	59
Tabel 4.3	Jumlah data Confussion Matrix	59

DAFTAR GAMBAR

Gambar 3.1	Alur Sistem	9
Gambar 3.2	Alur Preprocessing	21
Gambar 3.3	Wireframe Dashboard	34
Gambar 3.4	Wireframe dataset training	35
Gambar 3.5	Wireframe halaman bobot TF-IDF	35
Gambar 3.6	Wireframe halaman bobot Bayes	36
Gambar 3.7	Wireframe dataset testing	37
Gambar 3.8	Wireframe halaman preprocessing	37
Gambar 3.9	Wireframe halaman klasifikasi data	38
Gambar 4.1	Fungsi Crawling data	39
Gambar 4.2	Fungsi Case Folding	41
Gambar 4.3	Fungsi Cleansing	42
Gambar 4.4	Fungsi Konversi Emoji	43
Gambar 4.5	Fungsi Konversi Negasi	44
Gambar 4.6	Fungsi Normalisasi	45
Gambar 4.7	Fungsi Tokenisasi	46
Gambar 4.8	Fungsi stemming	46
Gambar 4.9	Fungsi Term Frequency (TF)	47
Gambar 4.10	Fungsi Inverse Document Frequency (IDF)	47
Gambar 4.11	Fungsi (TF-IDF)	48
Gambar 4.12	Fungsi Pembobotan Naive Bayes	49
Gambar 4.13	Fungsi Pembobotan Data Testing	50
Gambar 4.14	Fungsi Hitung Probabilitas Data Testing	51
Gambar 4.15	Fungsi Prediksi Sentimen Data Testing	53
Gambar 4.16	Halaman Dashboard	53
Gambar 4.17	Halaman Data Training	54
Gambar 4.18	Halaman Bobot TF-IDF	55
Gambar 4.19	Halaman Bobot NBC	55
Gambar 4.20	Halaman Data Testing	56

Gambar 4.21 **Halaman *Preprocessing***..... 57
Gambar 4.22 **Halaman Klasifikasi Data Testing**..... 58
Gambar 4.23 **Diagram Koleksi Data *Training***..... 58



DAFTAR ISTILAH

Bootstrap adalah sebuah *library framework* CSS yang di buat khusus untuk bagian pengembangan *front-end website*

Crawling adalah proses mengambil kumpulan data dari halaman sebuah web untuk dilakukan pengindeksan

Framework digambarkan sebagai abstraksi, di mana perangkat lunak yang menyediakan fungsionalitas generik dapat diubah secara selektif berdasarkan penggunaan kode penulisan.

Library digambarkan sebagai “sekumpulan implementasi dari tingkah laku, yang ditulis dalam bahasa pemrograman, memiliki antarmuka yang jelas saat perilaku itu dipanggil.

PHP adalah bahasa pemrograman *script server-side* yang didesain untuk pengembangan web.

Probabilitas adalah peluang atau kemungkinan dari suatu kejadian, terjadi, atau tidak dan seberapa besar kemungkinan kejadian tersebut berpeluang untuk terjadi

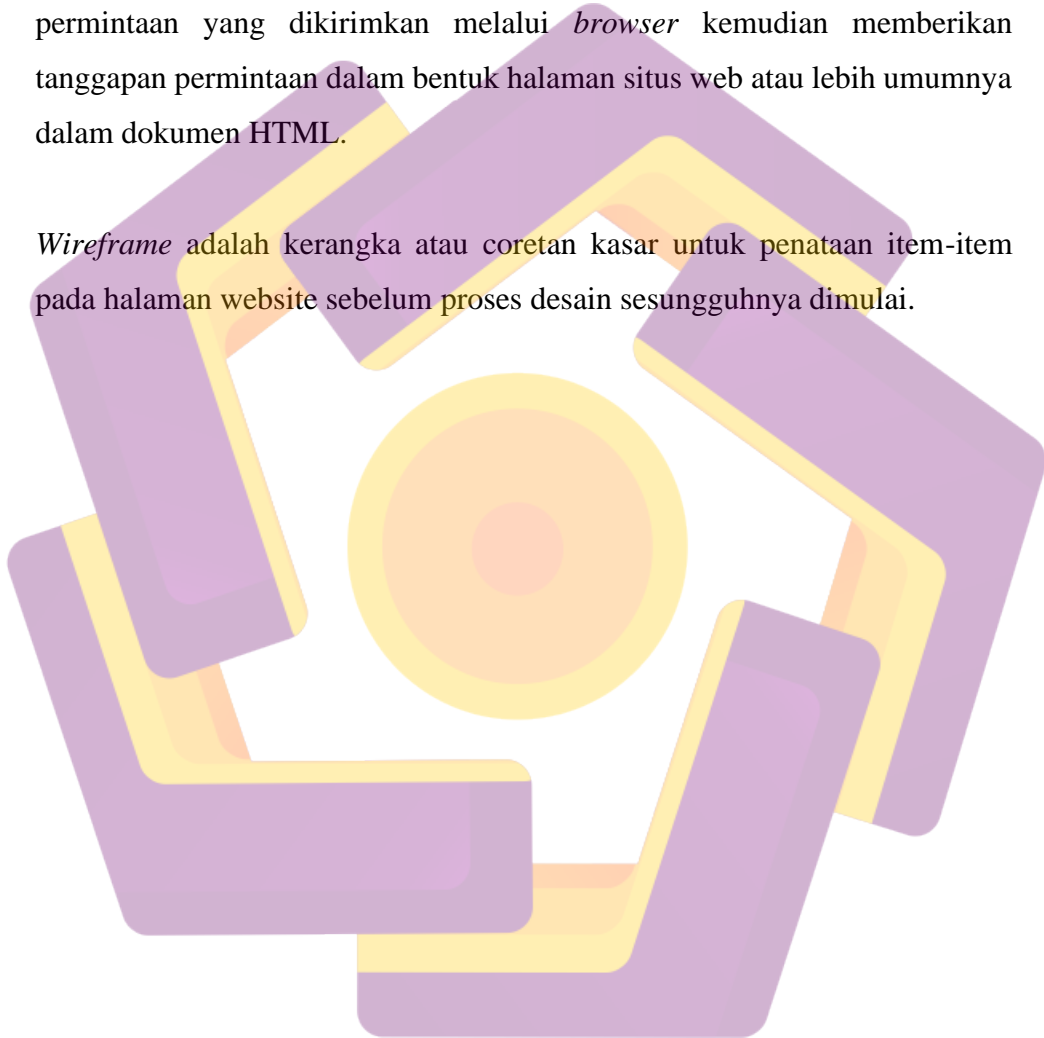
Sastrawi *stemmer* merupakan sebuah *library stemmer* sederhana yang didesain untuk dapat digunakan secara mudah.

Script merupakan suatu *code* yang digunakan untuk menerjemahkan setiap perintah dalam *website*.

Text Editor adalah suatu software aplikasi atau suatu program komputer yang memungkinkan penggunanya untuk membuat, mengubah atau mengedit file teks yang ada berupa *plain text*.

Web server adalah perangkat lunak yang berfungsi sebagai penerima permintaan yang dikirimkan melalui *browser* kemudian memberikan tanggapan permintaan dalam bentuk halaman situs web atau lebih umumnya dalam dokumen HTML.

Wireframe adalah kerangka atau coretan kasar untuk penataan item-item pada halaman website sebelum proses desain sesungguhnya dimulai.



INTISARI

Kata Kunci: Analisis Sentimen, Media Sosial, Klasifikasi, Naive Bayes Classifier, TF-IDF

Wisata di Indonesia merupakan salah satu penghasil devisa terbesar negara, pada tahun 2015 presiden Indonesia Joko Widodo memberikan amanat presiden melalui surat Sekretariat Kabinet Nomor B 652/Seskab/Maritim/2015 yang berisi arahan mengenai pariwisata untuk pengembangan 10 destinasi prioritas yang kemudian mulai berkembang pada tahun 2017 dan menjadi salah satu agenda prioritas bagi Menteri Pariwisata dan Ekonomi Kreatif periode 2019-2024 yaitu Wishnutama.

Dalam perkembangannya salah satu yang menjadi poin penting untuk mempromosikan 10 destinasi wisata prioritas adalah *social strategy* yaitu membangun hubungan antara pelaku pariwisata dengan wisatawan. *Social strategy* digunakan sebagai pendekatan terhadap wisatawan dengan memperhatikan interaksinya dengan orang lain di media sosial dengan target yaitu penjualan atau *brand awareness*. Dengan mengetahui sentiment, pendapat, keinginan masyarakat Indonesia terhadap tempat wisata, hal itu dapat digunakan sebagai acuan bagi pengelola maupun pemerintah untuk berbenah diri dan meningkatkan kualitas wisata agar memberikan layanan terbaik bagi pengunjung, atau menyusun strategi baru apabila memang dibutuhkan untuk memperkuat kualitas pariwisata di Indonesia.

Sifat media sosial yang terbuka membuat penggunaanya bebas mengekspresikan pendapatnya. Twitter merupakan salah satu sosial media yang digemari oleh masyarakat Indonesia. Penelitian ini mencoba memanfaatkan apa yang pengguna tulis atau posting di media sosial Twitter atau lebih dikenal dengan istilah *tweet*. *Tweet* tersebut nantinya akan diolah dengan *text mining* dan diproses menggunakan metode Naive Bayes Classifier.

ABSTRACT

Keyword: Sentiment Analysis, Social Media, Classification, Naive Bayes Classifier, TF-IDF

Tourism in Indonesia is one of the country's largest foreign exchange earners, in 2015 Indonesian president Joko Widodo gave the president's mandate through a Cabinet Secretariat letter B 652 / Seskab / Maritim / 2015 which contained directions on tourism for the development of 10 priority destinations which then began to develop in the year 2017 and become one of the priority agenda for the Minister of Tourism and Creative Economy 2019-2024 period, namely Wishnutama.

In its development, one of the important points to promote 10 priority tourist destinations is social strategy, namely building relationships between tourism actors and tourists. Social strategy is used as an approach to tourists by paying attention to their interactions with others on social media with a target of sales or brand awareness. By knowing the sentiments, opinions, desires of the Indonesian people towards tourist attractions, it can be used as a reference for managers and the government to improve themselves and improve the quality of tourism in order to provide the best service for visitors, or develop new strategies if needed to strengthen the quality of tourism in Indonesia. .

The open nature of social media makes users free to express their opinions. Twitter is a social media that is favored by the people of Indonesia. This research tries to utilize what users write or post on Twitter social media or better known as tweets. Tweets will be processed by text mining and processed using the Naive Bayes Classifier method.

