

***TEXT MINING* UNTUK ANALISIS SENTIMEN *REVIEW*
PARFUM HMNS MENGGUNAKAN ALGORITMA
NAÏVE BAYES**

SKRIPSI



disusun oleh

Ayu Hajar Nastiti Priyadi

17.12.0264

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM
YOGYAKARTA
2021**

TEXT MINING UNTUK ANALISIS SENTIMEN REVIEW
PARFUM HMNS MENGGUNAKAN ALGORITMA
NAÏVE BAYES

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Ayu Hajar Nastiti Priyadi

17.12.0264

PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM
YOGYAKARTA
2021

PERSETUJUAN

SKRIPSI

***TEXT MINING* UNTUK ANALISIS SENTIMEN *REVIEW* PARFUM HMNS MENGGUNAKAN ALGORITMA NAIVE BAYES**

yang dipersiapkan dan disusun oleh

Ayu Hajar Nastiti Priyadi

17.12.0264

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Agustus 2021

Dosen Pembimbing,

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412

PENGESAHAN

SKRIPSI

TEXT MINING UNTUK ANALISIS SENTIMEN *REVIEW* PARFUM HMNS MENGGUNAKAN ALGORITMA NAIVE BAYES

yang dipersiapkan dan disusun oleh

Ayu Hajar Nastiti Priyadi

17.12.0264

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Oktober 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Heri Sismoro, M.Kom.
NIK. 190302057

Yuli Astuti, M.Kom.
NIK. 190302146

Yoga Pristyanto, S.Kom.,M.Eng.
NIK. 190302412

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Oktober 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.
NIK. 190302096

PERNYATAAN

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, 2 November 2021



Ayu Hajar Nastiti Priyadi

NIM. 17.12.0264

MOTTO

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا ۗ لَهَا مَا كَسَبَتْ وَعَلَيْهَا مَا اكْتَسَبَتْ ۗ رَبَّنَا لَا تُؤَاخِذْنَا إِنْ نَسِينَا أَوْ أَهْطَأْنَا ۗ رَبَّنَا وَلَا تَحْمِلْ عَلَيْنَا إصْرًا كَمَا حَمَلْتَهُ عَلَى الَّذِينَ مِنْ قَبْلِنَا ۗ رَبَّنَا وَلَا تُحَمِّلْنَا مَا لَا طَاقَةَ لَنَا بِهِ ۗ وَاعْفُ عَنَّا وَارْحَمْنَا ۗ أَنْتَ مَوْلَانَا فَانصُرْنَا عَلَى الْقَوْمِ الْكَافِرِينَ

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Ia mendapat pahala (dari kebajikan) yang diusahakannya dan ia mendapat siksa (dari kejahatan) yang dikerjakannya. (Mereka berdoa): "Ya Tuhan kami, janganlah Engkau hukum kami jika kami lupa atau kami tersalah. Ya Tuhan kami, janganlah Engkau bebankan kepada kami beban yang berat sebagaimana Engkau bebankan kepada orang-orang sebelum kami. Ya Tuhan kami, janganlah Engkau pikulkan kepada kami apa yang tak sanggup kami memikulnya. Beri maafilah kami, ampunilah kami dan rahmatilah kami. Engkaulah Penolong kami, maka tolonglah kami terhadap kaum yang kafir." (QS. Al-Baqarah: 286).

"Everything will be okay in the end. If it is not okay, it's not the end"

- John Lennon

PERSEMBAHAN

Puji Syukur kepada Allah SWT yang telah memberikan limpahan rahmat dan hidayah sehingga skripsi ini dapat terselesaikan. Tidak lupa sholawat dan salam penulis haturkan kepada junjungan besar kita yaitu Nabi Muhammad SAW.

Pada halaman persembahan ini penulis ingin menyampaikan rasa syukur dan terima kasih kepada :

- Allah SWT yang telah memberikan rahmat berupa kesehatan, kekuatan, kesabaran, ketekunan, fokus, dan semangat sehingga seluruh rangkaian proses pengerjaan skripsi dapat berjalan dengan baik hingga skripsi ini selesai serta memperoleh hasil yang maksimal.
- Orang Tua (Bapak Drs. Supriyadi & Ibu Maryati, S.Pd.SD) yang senantiasa mendoakan dan memberi dukungan.
- Bapak Yoga Pristyanto, S.Kom.,M.Eng. selaku dosen pembimbing yang telah membimbing dan memberi arahan sehingga proses pengerjaan skripsi dapat dilakukan dengan lancar.
- Seluruh dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu selama perkuliahan.
- Teman-teman kelas Sistem Informasi angkatan 2017 khususnya kelas SI-04 yang telah menemani berjuang bersama-sama selama perkuliahan.

KATA PENGANTAR

Puji syukur penulis haturkan kepada Allah SWT Tuhan Yang Maha Esa atas segala rahmat dan berkat-Nya penulis dapat menyusun skripsi ini dengan baik dan lancar.

Perlu diketahui bahwa skripsi ini merupakan salah satu syarat kelulusan bagi setiap mahasiswa program studi Sistem Informasi Universitas AMIKOM Yogyakarta, serta merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program Strata-1 dan untuk memperoleh gelar Sarjana Komputer.

Dengan selesainya skripsi ini, maka penulis tidak lupa mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Anggit Dwi Hartanto, M.Kom. selaku ketua program studi Sistem Informasi UNIVERSITAS AMIKOM Yogyakarta.
3. Bapak Yoga Pristyanto, S.Kom.,M.Eng. selaku dosen pembimbing yang telah membimbing dan memberi arahan sehingga proses pengerjaan skripsi dapat dilakukan dengan lancar.
4. Seluruh dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu selama perkuliahan.
5. Seluruh teman-teman kuliah terkhusus untuk keluarga besar S1 Sistem Informasi 04 yang telah menemani dan memberikan pengalaman serta kenangan indah selama perkuliahan.

Penulis menyadari bahwa penyusunan skripsi ini masih jauh dari sempurna, dikarenakan keterbatasan pengetahuan dan pengalaman penulis. Serta dengan terbuka mengucapkan terima kasih atas kritik dan saran dari pembaca guna perbaikan pada masa mendatang.

Yogyakarta, 2 November 2021

Ayu Hajar Nastiti Priyadi

17.12.0264

DAFTAR ISI

JUDUL	I
PERSETUJUAN.....	II
PENGESAHAN	III
PERNYATAAN.....	IV
MOTTO.....	V
PERSEMBAHAN	VI
KATA PENGANTAR.....	VII
DAFTAR ISI	VIII
DAFTAR TABEL	X
DAFTAR GAMBAR.....	XI
DAFTAR ISTILAH.....	XII
INTISARI.....	XIII
<i>ABSTRACT</i>	XIV
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	3
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN.....	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Analisis Permasalahan	5
1.6.3 Implementasi	5
1.6.4 Analisis Hasil.....	5
1.7 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI	7
2.1 KAJIAN PUSTAKA.....	7

2.2	DASAR TEORI.....	12
2.2.1	Text Mining.....	12
2.2.2	Twitter & Twitter API.....	12
2.2.3	Crawling Data.....	16
2.2.4	Python.....	16
2.2.5	Klasifikasi.....	18
2.2.6	Analisis Sentimen.....	18
2.2.7	Preprocessing.....	19
2.2.8	Naive Bayes Classifier.....	21
2.2.9	Term Frequency – Inverse Document Frequency (TF-IDF).....	22
2.2.10	K-Fold Cross Validation.....	24
2.2.11	Confusion Matrix.....	25
BAB III METODOLOGI PENELITIAN.....		28
3.1	METODE PENGUMPULAN DATA.....	28
3.1.1	Studi Pustaka.....	28
3.1.2	Studi Literatur Sejenis.....	28
3.2	ALUR PENELITIAN.....	28
BAB IV HASIL DAN PEMBAHASAN.....		49
4.1	RANCANGAN SISTEM.....	49
4.2	PREPROCESSING.....	52
4.3	PEMBOBOTAN KATA.....	60
4.4	IMPLEMENTASI KLASIFIKASI NAÏVE BAYES.....	67
4.5	UJI MODEL.....	70
4.6	EVALUASI MODEL.....	71
BAB V PENUTUP.....		76
5.1	KESIMPULAN.....	76
5.2	SARAN.....	77
DAFTAR PUSTAKA.....		78

DAFTAR TABEL

<u>Tabel 2.1</u>	Perbandingan penelitian terkait	9
<u>Tabel 2.2</u>	contoh <i>confussion matrix</i> dua kelas.....	25
<u>Tabel 3.1</u>	Contoh data tweet	29
<u>Tabel 3.2</u>	Contoh hasil <i>cleansing</i>	30
<u>Tabel 3.3</u>	Contoh hasil <i>case folding</i>	32
<u>Tabel 3.4</u>	Contoh hasil <i>tokenizing</i>	33
<u>Tabel 3.5</u>	Contoh hasil normalisasi	35
<u>Tabel 3.6</u>	Contoh hasil <i>stopword removal</i>	36
<u>Tabel 3.7</u>	Contoh hasil <i>stemming</i>	37
<u>Tabel 3.8</u>	Perhitungan TF	39
<u>Tabel 3.9</u>	Perhitungan DF	40
<u>Tabel 3.10</u>	Perhitungan IDF	41
<u>Tabel 3.11</u>	Perhitungan TF-IDF	42
<u>Tabel 3.12</u>	Contoh data latih.....	43
<u>Tabel 3.13</u>	Contoh data uji.....	44
<u>Tabel 3.14</u>	<i>Confussion Matrix</i>	47
<u>Tabel 4.1</u>	Data Bersih	61
<u>Tabel 4.2</u>	<i>Word Vector</i>	62
<u>Tabel 4.3</u>	<i>Term Frequency</i>	63
<u>Tabel 4.4</u>	<i>Document Frequency</i>	63
<u>Tabel 4.5</u>	<i>Inverse Document Frequency</i>	64
<u>Tabel 4.6</u>	TF-IDF	65
<u>Tabel 4.7</u>	Bobot	66
<u>Tabel 4.8</u>	Confusion Matrix Kelas A-B.....	70
<u>Tabel 4.9</u>	Hasil <i>Confusion Matrix</i> NB.....	71
<u>Tabel 4.10</u>	Hasil Evaluasi NB	74
<u>Tabel 4.11</u>	Hasil Evaluasi <i>K-Fold Cross Validation</i>	75

DAFTAR GAMBAR

<u>Gambar 2.1</u>	Ilustrasi <i>K-Fold Cross Validation</i>	24
<u>Gambar 3.1</u>	Tahapan proses sentimen analisis.....	29
<u>Gambar 3.2</u>	Kamus Normalisasi	34
<u>Gambar 4.1</u>	<i>API Key Twitter</i>	49
<u>Gambar 4.2</u>	Kode Proses <i>Crawling</i>	50
<u>Gambar 4.3</u>	Hasil <i>crawling</i>	51
<u>Gambar 4.4</u>	<i>Labeling</i>	51
<u>Gambar 4.5</u>	Presentase Data Hasil <i>Labeling</i>	52
<u>Gambar 4.6</u>	<i>Cleansing</i>	53
<u>Gambar 4.7</u>	<i>Case folding</i>	54
<u>Gambar 4.8</u>	<i>Tokenizing</i>	55
<u>Gambar 4.9</u>	Kamus <i>Slang Word</i>	56
<u>Gambar 4.10</u>	Normalisasi.....	56
<u>Gambar 4.11</u>	<i>Stopword Removal</i>	58
<u>Gambar 4.12</u>	<i>Stemming</i>	59
<u>Gambar 4.13</u>	Visualisasi <i>Word Cloud</i>	61
<u>Gambar 4.14</u>	<i>Count Vectorizer</i> dan TF-IDF	67
<u>Gambar 4.15</u>	Deklarasi Lib NB.....	68
<u>Gambar 4.16</u>	Pemanggilan Data.....	68
<u>Gambar 4.17</u>	Klasifikasi NB	69
<u>Gambar 4.18</u>	<i>Confusion Matrix</i>	71
<u>Gambar 4.19</u>	Hasil Evaluasi.....	73
<u>Gambar 4.20</u>	Hasil <i>K-Fold Cross Validation</i>	75

INTISARI

Pengguna media sosial saat ini sangat besar, dimana setiap orang mengungkapkan pendapat, komentar, kritik dan lain-lain di dalamnya. Salah satu media sosial yang paling banyak digunakan yaitu *twitter*.

Twitter seringkali digunakan untuk mengungkapkan ulasan atau *review* terhadap suatu produk. Data tersebut memberikan informasi yang dapat membantu orang atau organisasi dalam pengambilan keputusan, karena didalam sebuah ulasan terdapat informasi dalam bentuk emosi yang dapat diketahui sentimennya apakah negatif atau positif. Namun, jumlah data yang sangat besar tidak mungkin bagi manusia untuk membaca dan menganalisis secara manual.

Analisis Sentimen merupakan proses dalam menganalisa pendapat, sentimen, evaluasi, penilaian dan emosi seseorang terhadap suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Penelitian ini dilakukan untuk mengetahui performa algoritma *Naive Bayes Classifier* dalam melakukan proses klasifikasi berdasarkan *review* atau ulasan konsumen sebuah produk parfum yaitu HMNS yang terdapat pada *twitter*. Metode *Naive Bayes Classifier* dan seleksi fitur *Term-Frequency* diterapkan pada penelitian ini. Dari penelitian ini diperoleh hasil akurasi Algoritma *Naive Bayes Classifier* dengan seleksi fitur *Term-Frequency* sebesar 86%, dilakukan juga pengujian menggunakan *k-fold cross validation* dengan $k = 10$ menghasilkan akurasi sebesar 82%.

Kata Kunci: *Review, Twitter, Analisis Sentimen, Naive Bayes Classifier, K-Fold Cross Validation*

ABSTRACT

Today's social media users are very large, where everyone expresses opinions, comments, criticisms and others in it. One of the most widely used social media is Twitter.

Twitter is often used to express reviews or reviews of a product. The data provides information that can help people or organizations in making decisions, because in a review there is information in the form of emotions that can be known whether the sentiment is negative or positive. However, the huge amount of data is impossible for humans to read and analyze manually.

Sentiment analysis is the process of analyzing one's opinions, sentiments, evaluations, judgments and emotions on a particular topic, product, service, organization, individual, or activity. This research was conducted to determine the performance of the Naïve Bayes Classifier algorithm in conducting the classification process based on consumer reviews of a perfume product, namely HMNS on Twitter. The Naïve Bayes Classifier method and Term-Frequency feature selection are applied in this study. From this study, the results of the accuracy of the Naïve Bayes Classifier Algorithm with the Term-Frequency feature selection of 86% were also carried out using k-fold cross validation with $k = 10$ resulting in an accuracy of 82%.

Keyword: *Review, Twitter, Sentiment Analysis, Naïve Bayes Classifier, K-Fold Cross Validation*