

ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA MENGGUNAKAN NAIVE BAYES

SKRIPSI

**Diajukan untuk memenuhi persyaratan
dalam memperoleh gelar Sarjana Komputer
Program Studi Sistem Informasi
Disusun Oleh:**



AISYAH FIRDAUSI RAHMA

17082010055

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
S U R A B A Y A
2021**

SKRIPSI

ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA MENGUNAKAN NAIVE BAYES

Disusun Oleh:

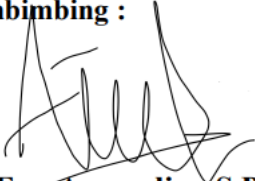
AISYAH FIRDAUSI RAHMA

17082010055

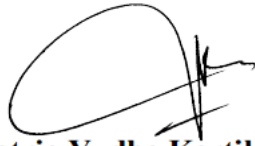
Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi
Program Studi Sistem Informasi Fakultas Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jawa Timur
Pada Tanggal 8 Januari 2021

Pembimbing :

1.

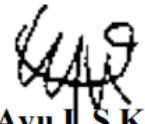

Dr. Eng. Agussalim, S.Pd., M.T.
NIP. 19850811 20190310 05

2.

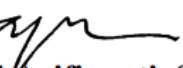

Dhian Satria Yudha Kartika, S.Kom, M.Kom.
NPT. 201198 60 522249

Tim Penguji:

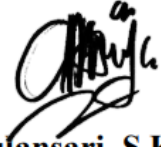
1.


Syurfah Ayu I., S.Kom, M.Kom.
NPT. 3 8501 10 0294 1

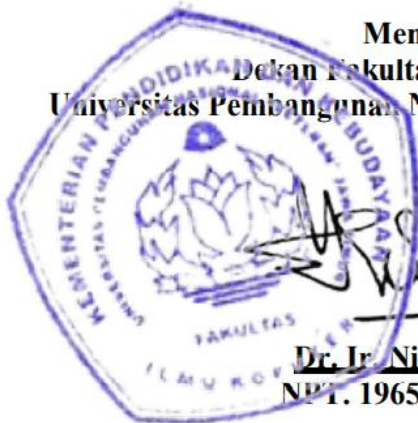
2.


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

3.


Anita Wulansari, S.Kom., M.Kom.
NPT. 202198 71 015202

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



Dr. Ir. Ni Ketut Sari, MT.
NPT. 19650731199203 2 001

LEMBAR PENGESAHAN

**ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA
MENGUNAKAN NAIVE BAYES**

Disusun Oleh:

AISYAH FIRDAUSI RAHMA

17082010055

**Telah disetujui mengikuti Ujian negara Lisan Gelombang Januari
Periode 2021 pada Tanggal 8 Januari 2021**

Dosen Pembimbing 1



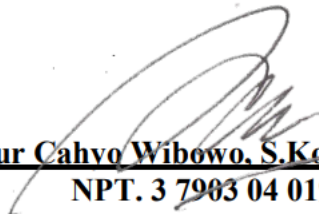
Dr. Eng. Agussalim, S.Pd., M.T.
NIP. 19850811 20190310 05

Dosen Pembimbing 2



Dhian Satria Yudha Kartika, S.Kom, M.Kom.
NPT. 201198 60 522249

**Mengetahui,
Ketua Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jawa Timur**



Nur Cahyo Wibowo, S.Kom, M.Kom
NPT. 3 7903 04 0197 1

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR
FAKULTAS ILMU KOMPUER**

KETERANGAN REVISI

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : Aisyah Firdausi Rahma

NPM : 17082010055

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 8 Januari 2021 dengan judul:

**ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA
MENGUNAKAN NAIVE BAYES**

Oleh karenanya mahasiswa tersebut diatas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, Januari 2021

Dosen penguji yang memeriksa revisi:

1. Syurfah Ayu I, S.Kom, M.Kom.
NPT. 3 8501 10 0294 1
2. Amalia Anjani Arifiyanti, S.Kom, M.Kom.
NIP. 19920812 201803 2 001
3. Anita Wulansari, S.Kom., M.Kom.
NPT. 202198 71 015202

Mengetahui,



Dosen Pembimbing 1



Dr.Eng Agussalim, S.Pd., M.T.
NIP. 19850811 20190310 05

Dosen Pembimbing 2



Dhian Satria Yudha Kartika, S.Kom, M.Kom.
NPT. 201198 60 522249



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA
TIMUR
FAKULTAS ILMU KOMPUTER**

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : AISYAH FIRDAUSI RAHMA

NPM : 17082010055

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA
MENGUNAKAN NAIVE BAYES**

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / Software / Hasil Karya yang saya beli dari 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 di Institusi Pendidikan lain. Jika ternyata dikemudian hari pernyataan ini terbukti benar, maka Saya bertanggungjawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah di kemudian hari.

Surabaya, 19 Januari 2021



Hormat saya,

AISYAH FIRDAUSI RAHMA

17082010055

Judul : Analisis Sentimen Hashtag Kuliner di Indonesia
Menggunakan Naive Bayes
Pembimbing 1 : Dr. Eng. Agussalim, M.T.
Pembimbing 2 : Dhian Satria Yudha Kartika, S.Kom, M.Kom

ABSTRAK

Seiring dengan berkembangnya zaman, teknologi baru semakin banyak bermunculan. Banyaknya teknologi baru tersebut mengakibatkan jumlah data yang dihasilkan semakin meningkat. Munculnya media sosial bernama twitter pada tahun 2006 juga ikut menyumbang peningkatan jumlah data yang ada. Penggunaan media sosial twitter ini dapat menjadikan penggunaannya berbagi tentang keseharian ataupun informasi dengan rekan atau kerabat secara *realtime*. Jumlah pengguna twitter yang sangat tinggi dapat menjadi sarana bagi masyarakat untuk melakukan jual beli (promosi), dapat juga untuk menuliskan perasaan apa yang dirasakan, produk apa saja yang sedang digunakan dan juga referensi tempat makan atau kuliner di Indonesia. Hal tersebut adalah salah satu alasan mengapa twitter merupakan media sosial yang tepat didalam melakukan analisis sentimen. Analisis sentimen sendiri perlu dilakukan karena penggunaan media sosial yang semakin meningkat sehingga dapat mempengaruhi perkembangan opini publik.

Pembuatan model analisis sentimen dengan *hashtag* kuliner pada twitter di Indonesia ini dibangun menggunakan python. Penelitian dilakukan dengan melakukan tahapan pengumpulan data, *preprocessing*, pelabelan data (positif, negatif, dan netral), pembobotan teks, pembagian data, Naive Bayes *Classifier*, evaluasi model klasifikasi serta visualisasi.

Model ini dibangun dengan menggunakan dua skenario. Pada tahapan pembobotan teks, skenario pertama adalah menggunakan *Bag of Words* selanjutnya pada skenario kedua menggunakan *TF-IDF*. Dataset yang digunakan memiliki jumlah 748 *tweet* yang diambil dalam waktu 1 Januari 2020 hingga 30 November 2020. Dataset sudah diberi 3 label yaitu positif, negatif dan netral. Pengambilan dataset diambil dengan menggunakan kata kunci yang memiliki *hashtag* kuliner. Hasil pemodelan klasifikasi terbaik adalah dengan menggunakan Multinomial Naive Bayes pada skenario 1 dengan nilai akurasi 0.71, rata-rata *precision* 0.80, rata-rata *recall* 0.53, dan rata-rata *f-score* 0.53.

Kata kunci: analisis sentimen, klasifikasi, Twitter, Naive Bayes

KATA PENGANTAR

Dengan mengucapkan syukur Alhamdulillah kepada Allah SWT, yang telah melimpahkan berkah, rahmat serta hidayah-Nya, sehingga dapat menyelesaikan skripsi ini dengan judul “ANALISIS SENTIMEN HASHTAG KULINER DI INDONESIA MENGGUNAKAN NAIVE BAYES”. Skripsi ini disusun sebagai salah satu persyaratan untuk memperoleh gelar Sarjana (S-1) Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Terselesaikannya skripsi ini tentunya tak lepas dari dorongan uluran tangan dari berbagai pihak. Oleh karena itu penulis mengucapkan rasa terima kasih dan penghargaan kepada :

- 1) Mama Puspita Kusuma Wardhani dan Ayah Basuki Rahmat sebagai orangtua yang senantiasa memberikan do’a, semangat, pendengar yang baik, asupan makanan, pengingat ibadah selama pembuatan laporan skripsi ini.
- 2) Hafshah Sakinah dan Hanin Azka Zahira, adik-adik ku tercinta yang suka ganggu tapi juga selalu mendo’akan dan juga kadang-kadang memijat kakaknya.
- 3) Bapak Dr.Eng Agussalim, S.Pd., M.T. Selaku dosen Pembimbing 1 yang senantiasa memberikan bimbingan, dorongan dan arahan dalam penyelesaian skripsi.
- 4) Bapak Dhian Satria Yudha Kartika, S.Kom, M.Kom. Selaku dosen Pembimbing 2 yang senantiasa memberikan bimbingan, dorongan dan arahan dalam penyelesaian skripsi.
- 5) Terimakasih kepada Ibu Eka sebagai dosen pengampu Data Mining, penguji sempro rasa pembimbing, Terimakasih banyak bu. Terimakasih juga kepada Ibu Amel, beliau adalah dosen penguji semhas akan tetapi seperti dosen pembimbing juga! Terimakasih banyak bu!!
- 6) Seluruh keluarga saya dari Eyang Muji dan Eyang Rosyid yang telah memberikan bantuan berupa do’a dan dukungan dalam penyelesaian skripsi ini.
- 7) Teman-teman seperjuangan yang sudah lulus maupun masih dalam proses pengerjaan skripsi yaitu Aya Maharani, Ersalina Adani, Dinda Siti, Annisa

Wibowo, Nuke Nidya, Shania Virgiani, Fari Handhina, Tentra Olivia, Adelia Sefri, Nurfian Triantono dan lain sebagainya.

- 8) Seluruh teman-teman Grup Hunting, ayo kita hunting lagi.
- 9) Teman-teman HIMASIFO “GELORA INTEGRITAS” dan BEM FASILKOM “KABINET SEROJA” terimakasih sudah mengizinkan untuk berproses bersama.
- 10) Ucapan terimakasih kepada diri saya sendiri. “Terimakasih kamu hebat!, sudah berjuang sampai tahap ini. Ayo semangat! Jalanmu masih panjang! Jangan Cuma kejar dunia, akhirat tetap nomor 1”.
- 11) Seluruh teman-teman INSYDIOUS, meskipun di dua semester terakhir kita terpisahkan oleh corona. Semoga semakin sukses dunia akhirat!!!

Surabaya, Januari 2021

Penulis

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan.....	6
1.5 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Dasar Teori	8
2.1.1 Twitter.....	8
2.1.2 Analisis Sentimen.....	8
2.1.3 <i>Crawling</i>	9
2.1.4 <i>Rule Based Sentiment</i>	9
2.1.5 <i>Text Preprocessing</i>	10
2.1.6 <i>Bag of Words</i>	10
2.1.7 <i>Term Frequency – Inverse Document Frequency</i>	11
2.1.8 Naive Bayes	12
2.1.9 Evaluasi Model.....	14
2.1.10 Bahasa Pemrograman Python	15
2.2 Penelitian Terdahulu.....	15
2.2.1 Hasil Penelitian Nabila Surya Wardani, Alan Prahutama, Puspita Kartikasari 2020	16
2.2.2 Hasil Penelitian Retno Sari, Ratih Yulia Hayuningtyas 2019.....	16
2.2.3 Hasil Penelitian Lingga Aji, Pratiwi Amalia, Respatiwan 2019..	17
2.2.4 Hasil Penelitian Mungki Astiningrum, Mamluatul Hani'ah, Yanuar Rahmat Yoga Pradana 2020.....	18
2.2.5 Hasil Penelitian Bong Hyun Back dan Il Kyu Ha 2019.....	18
BAB III METODOLOGI PENELITIAN.....	20
3.1 Pengumpulan Data.....	20
3.2 <i>Preprocessing</i>	21
3.3 Pelabelan Data	22
3.4 Pembobotan Teks	23
3.5 Pembagian Data.....	23
3.6 Naive Bayes <i>Classifier</i>	23
3.7 Evaluasi Model Klasifikasi.....	24
3.8 Visualisasi.....	25
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Pengumpulan Data.....	26
4.2 <i>Preprocessing</i>	28

4.3	Pelabelan Data	32
4.4	Pembobotan Teks.....	33
4.5	Pembagian Data.....	34
4.6	Naive Bayes <i>Classifier</i>	34
4.7	Evaluasi Model Klasifikasi.....	37
4.8	Visualisasi.....	48
BAB V PENUTUP.....		52
5.1	Kesimpulan.....	52
5.2	Saran.....	53
DAFTAR PUSTAKA		54
LAMPIRAN		61

DAFTAR TABEL

Tabel 2. 1 Contoh Hasil <i>Confusion Matrix</i> Prediksi Dua Kelas	14
Tabel 4. 1 Contoh <i>tweet</i> yang di dapatkan	27
Tabel 4. 2 Contoh <i>tweet</i> yang sudah diterjemahkan	27
Tabel 4. 3 Contoh <i>tweet</i> yang melalui <i>cleaning</i> data	29
Tabel 4. 4 Contoh <i>tweet</i> yang melalui tokenisasi.....	30
Tabel 4. 5 Contoh <i>tweet</i> yang melalui <i>removing stopwords</i>	31
Tabel 4. 6 Contoh <i>tweet</i> yang melalui <i>stemming</i>	32

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alir Metodologi	20
Gambar 3. 2 Diagram Alir <i>Preprocessing</i>	21
Gambar 4. 1 <i>Source Code</i> Pengumpulan Data	26
Gambar 4. 2 Hasil <i>tweet</i> yang telah dikumpulkan	26
Gambar 4. 3 <i>Source Code</i> Penghapusan Kolom	28
Gambar 4. 4 <i>Source Code</i> untuk menghapus url.....	29
Gambar 4. 5 <i>Source Code</i> untuk menghapus simbol, tanda baca	29
Gambar 4. 6 <i>Source Code</i> untuk tokenisasi	30
Gambar 4. 7 <i>Source Code</i> untuk <i>removing stopwords</i>	31
Gambar 4. 8 <i>Source Code</i> untuk <i>stemming</i>	32
Gambar 4. 9 <i>Source Code</i> Pelabelan Data	32
Gambar 4. 10 Hasil <i>tweet</i> yang telah dilabelkan.....	33
Gambar 4. 11 <i>Source Code Bag of Words</i>	33
Gambar 4. 12 Hasil dari <i>Bag of Words</i>	33
Gambar 4. 13 <i>Source Code</i> TF-IDF	34
Gambar 4. 14 Hasil dari TF-IDF	34
Gambar 4. 15 <i>Source Code</i> Pembagian Data	34
Gambar 4. 16 <i>Source code</i> dan hasil Bernoulli NB (skenario 1)	35
Gambar 4. 17 <i>Source code</i> dan hasil Multinomial NB (skenario 1)	35
Gambar 4. 18 <i>Source code</i> dan hasil Bernoulli NB (skenario 2)	36
Gambar 4. 19 <i>Source code</i> dan hasil Multinomial NB (skenario 2)	36
Gambar 4. 20 <i>Source code</i> dan hasil perbandingan (skenario 1)	37
Gambar 4. 21 <i>Source code</i> dan hasil perbandingan (skenario 2)	37
Gambar 4. 22 <i>Source code</i> dan hasil confusion matrix Bernoulli NB (skenario 1)	
.....	37
Gambar 4. 23 Cara membaca <i>confusion matrix</i> 3x3	38
Gambar 4. 24 <i>Source code</i> dan hasil <i>confusion matrix</i> Multinomial NB (skenario	
2)	39
Gambar 4. 25 <i>Source code</i> dan hasil <i>confusion matrix</i> Bernoulli NB (skenario 2)	
.....	40
Gambar 4. 26 <i>Source code</i> dan hasil <i>confusion matrix</i> Multinomial NB (skenario	
2)	42
Gambar 4. 27 <i>Source code</i> dan hasil dari pengukuran <i>performance metrics</i> dari	
<i>confusion matrix</i> Bernoulli Naive Bayes pada skenario 1	43
Gambar 4. 28 <i>Source code</i> dan hasil dari pengukuran <i>performance metrics</i> dari	
<i>confusion matrix</i> Multinomial Naive Bayes pada skenario 1	45
Gambar 4. 29 <i>Source code</i> dan hasil dari pengukuran <i>performance metrics</i> dari	
<i>confusion matrix</i> Bernoulli Naive Bayes pada skenario 2	46
Gambar 4. 30 <i>Source code</i> dan hasil dari pengukuran <i>performance metrics</i> dari	
<i>confusion matrix</i> Multinomial Naive Bayes pada skenario 2	47
Gambar 4. 31 Grafik <i>polarity</i> keseluruhan	48
Gambar 4. 32 Jumlah <i>sentiment</i> keseluruhan.....	48
Gambar 4. 33 Presentase <i>sentiment</i>	49
Gambar 4. 34 Rata-rata <i>sentiment tweet</i> berdasarkan tanggal	49
Gambar 4. 35 <i>Wordcloud</i> dengan <i>polarity positive</i>	50
Gambar 4. 36 <i>Wordcloud</i> dengan <i>polarity negatif</i>	50

Gambar 4. 37 <i>Wordcloud</i> dengan <i>polarity neutral</i>	50
Gambar 4. 38 Daftar kuliner yang ramai diperbincangkan sesuai <i>polarity</i>	51

DAFTAR LAMPIRAN