

**DETEKSI EMOSI PADA ULASAN APLIKASI E-
TICKETING PADA SITUS GOOGLE PLAY
MENGUNAKAN METODE NAÏVE BAYES
(Studi kasus Aplikasi Traveloka)**

SKRIPSI

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



Disusun Oleh:

**ANA KHOLIFATUL JANAH
NPM 1635010022**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR
S U R A B A Y A
2020**

SKRIPSI

DETEKSI EMOSI PADA ULASAN APLIKASI E - TICKETING PADA SITUS GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES (Studi kasus Aplikasi Traveloka)

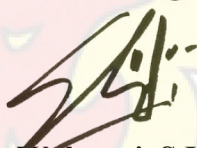
Disusun Oleh:

ANA KHOLIFATUL JANAH
NPM 1635010022

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 20 Juli 2020

Dosen Pembimbing :

1.

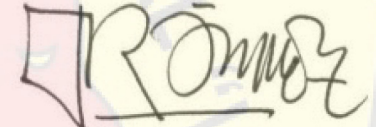

Eka Dyar Wahyuni, S.Kom., M.Kom
NPT. 3 8412 13 0356 1

2.


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

Dosen Penunji :

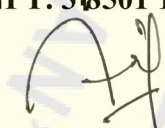
1.


M. Irwan Afandi, ST.M.Sc
NPT. 3 7607 07 0220 1

2.


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

3.


Asif Faruqi, S.Kom., M.Kom
NIP. 19870519 201803 1 001

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



Dr. Ir. I Ketut Sari, M.T.
NIP. 19650731 199203 2 001

LEMBAR PENGESAHAN
DETEKSI EMOSI PADA ULASAN APLIKASI E - TICKETING
PADA SITUS GOOGLE PLAY MENGGUNAKAN METODE
NAÏVE BAYES (Studi kasus Aplikasi Traveloka)

Disusun Oleh:

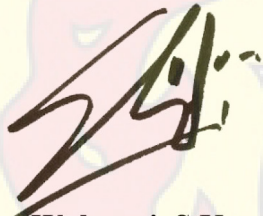
ANA KHOLIFATUL JANAH
NPM 1635010022

Telah disetujui mengikuti Ujian Negara Lisan Periode 2020 pada Tanggal 20
Juli 2020

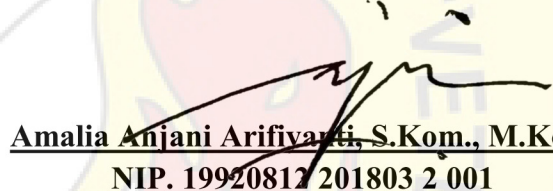
Menyetujui,

Dosen Pembimbing 1

Dosen Pembimbing 2

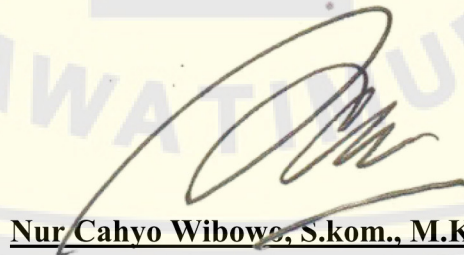


Eka Dyar Wahyuni, S.Kom., M.Kom.
NPT. 3 8412 13 0356 1



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

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



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

Jl. Rungkut Madya Gunung Anyar Surabaya 60294
Telp (031) 8706369, 8783189 Fax (031) 8706372 Website www.upnjatim.ac.id

KETERANGAN REVISI

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

Nama : Ana Kholifatul Janah

NPM : 1635010022

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 20 Juli 2020 dengan judul :

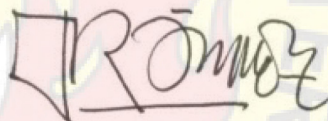
**DETEKSI EMOSI PADA ULASAN APLIKASI E - TICKETING PADA SITUS
GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES (Studi kasus
Aplikasi Traveloka)**

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

Dosen Penguji yang memeriksa revisi :

Surabaya, 20 Juli 2020

1. M. Irwan Afandi, ST.M.Sc
NPT. 3 7607 07 0220 1

{  }

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

{  }

3. Asif Faroqi, S.Kom., M.Kom
NIP. 19870519 201803 1 001

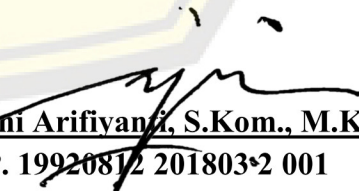
{  }

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2


Eka Dyar Wahyuni, S.Kom., M.Kom
NPT. 3 8412 13 0356 1


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



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

Jl. Rungkut Madya Gunung Anyar Surabaya 60294
Telp (031) 8706369, 8783189 Fax (031) 8706372 Website www.upnjatim.ac.id

SURAT PERNYATAAN

Saya, Mahasiswa Sistem Informasi UPN "Veteran" Jawa Timur, yang bertanda tangan di bawah ini :

Nama : Ana Kholifatul Janah

NPM : 1635010022

Program Studi : Sistem Informasi

Fakultas : Ilmu Komputer

Menyatakan Bahwa Judul Skripsi / Tugas Akhir Saya Sebagai Berikut :

**DETEKSI EMOSI PADA ULASAN APLIKASI E - TICKETING PADA SITUS
GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES (Studi kasus
Aplikasi Traveloka)**

Bukan merupakan plagiat dari skripsi / tugas akhir / penelitian orang lain dan juga bukan merupakan produk / software / hasil karya yang saya beli dari pihak 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 UPN "Veteran" Jawa Timur maupun institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka saya bertanggung jawab penuh dan siap menerima segala konsekuensinya, termasuk pembatalan ijazah dikemudian hari.

Hormat Saya,



Ana Kholifatul Janah

Judul Skripsi : DETEKSI EMOSI PADA ULASAN APLIKASI E-TICKETING PADA SITS GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Aplikasi Traveloka)

Penulis : Ana Kholifatul Janah

Pembimbing I : Eka Dyar Wahyuni, S.Kom, M.Kom

Pembimbing II : Amalia Anjani Arifiyanti, S.Kom, M.Kom

ABSTRAK

Pola konsumsi masyarakat yang semakin banyak menggunakan platform digital untuk memenuhi kebutuhan konsumsi masyarakat turut mempengaruhi industry travel. Sector online travel ini menyumbang lebih dari 38 persen dari perkiraan pertumbuhan ekonomi internet di Asia Tenggara. Traveloka telah disediakan di situs *Google Play* dan diunduh lebih dari 30 juta kali, menjadikannya aplikasi mobile paling populer di Asia Tenggara. Hal yang menarik dari aplikasi *google play* ialah adanya sebuah fitur yang dapat berisi ulasan dari para penggunanya yang bisa dianalisis untuk mengetahui emosi yang terkandung didalamnya.

Deteksi emosi yang dianalisis yakni berupa emosi senang, sedih, terkejut dan marah. Pada studi kasus ini, klasifikasi data yang dilakukan menggunakan metode Naïve Bayes yang digunakan ada 3 macam (Multinomial NB, Bernoulli NB, dan Gaussian NB), dengan tujuan untuk mendapatkan nilai akurasi yang terbaik dan memiliki sebaran nilai F1 score yang merata.

Dari penelitian yang dilakukan data ulasan aplikasi Traveloka yang didapat dari bulan November 2019 hingga bulan Januari 2020 yakni 2.662 ulasan, hasil akurasi terbaik didapatkan dengan menggunakan metode Multinomial NB dengan menggunakan metode *Hold Out* sebanyak 20% data test dengan akurasi sebesar 86% sedangkan metode Bernoulli NB akurasinya sebesar 69% dan Gaussian NB akurasinya sebesar 44%.

Kata Kunci: *Data Mining, Emosi, Google Play, Klasifikasi, Naïve Bayes, Traveloka.*

KATA PENGANTAR

Alhamdulillah *robbil 'alamin*, segala puji bagi Allah Yang Maha Kuasa yang telah memberikan kekuatan-Nya, serta sholawat dan salam tercurahkan pada junjungan Nabi Besar Muhammad SAW, sehingga skripsi berjudul Deteksi Emosi Pada Ulasan Aplikasi Pada Situs Google Play Menggunakan Metode Naïve Bayes yang merupakan persyaratan dalam menyelesaikan Program Studi Sistem Informasi di Universitas Pembangunan Nasional “Veteran” Jawa Timur ini dapat diselesaikan tepat waktu.

Melalui skripsi ini, saya merasa mendapat kesempatan besar untuk lebih memperdalam ilmu pengetahuan yang diperoleh selama di perkuliahan, terutama berkenaan dengan implementasi sistem informasi dalam kehidupan sehari-hari. Namun demikian, saya menyadari bahwa skripsi ini masih memiliki banyak kelemahan dan kekurangan.

Secara khusus, dalam kesempatan ini pula, ingin mengucapkan terima kasih dan penghargaan sebesar-besarnya kepada:

1. Bapak Ahmad Rifai, Ibu Juhairiyah dan kakak saya selaku keluarga yang telah memberikan dukungan materi dan doa agar skripsi ini bisa selesai serta mendoakan saya agar selalu menjadi orang yang bisa bermanfaat untuk lingkungan sekitar.
2. Ibu Eka Dyar Wahyuni, S.Kom, M.Kom selaku dosen pembimbing I yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi.

3. Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom selaku dosen pembimbing II yang selalu memberikan waktu dan arahan dalam menyelesaikan skripsi.
4. Seluruh dosen Program Studi Sistem Informasi atas ilmu-ilmu yang di berikan selama masa perkuliahan.
5. Teman-teman Rosa, Arilza, Kiki, Lisa dan Afifatul yang selama ini selalu menemani saya dalam suka dan duka.
6. Ega Febri Dharmawan, S. Kom selaku alumni pembimbing.
7. Teman-teman Program Studi Sistem Informasi yang telah membantu saya dalam menyelesaikan skripsi ini.

Terima kasih kepada semua pihak yang telah membantu penyelesaian skripsi ini. Semoga Allah membalas dengan balasan sebaik-baiknya. Dalam pengerjaan skripsi ini saya menyadari bahwa skripsi ini masih ada kekurangan.

Sebagai manusia biasa pasti mempunyai keterbatasan dan banyak sekali kekurangan, terutama dalam pembuatan skripsi ini. Untuk itu saya sangat membutuhkan kritik dan saran yang membangun dalam memperbaiki penulisan skripsi ini.

Surabaya

Penulis

DAFTAR ISI

LEMBAR PERNYATAAN.....	ii
LEMBAR PENGESAHAN.....	iii
KETERANGAN REVISI.....	iv
SURAT PERYATAAN ANTI PLAGIASI.....	v
ABSTRAK.....	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan	6
1.5 Manfaat	6
1.6 Sistematika Penulisan	6
BAB II .TINJAUAN PUSTAKA.....	7
2.1 Landasan Teori.....,.....	7
2.1.1 Explatory Analysis Data (EDA).....	7
2.1.2 Emosi	8

2.1.3 Text Mining.....	9
2.1.4 Analisis Sentimen.....	9
2.1.5 Term frequency – inverse document frequency (TF-IDF).....	9
2.1.6 Naïve Bayes.....	10
2.1.7 Evaluasi Performa Model.....	12
2.1.8 Google Play.....	13
2.1.9 E-Ticketing.....	14
2.1.10 Traveloka.....	14
2.1.11 Python.....	14
2.1.12 Penelitian Terdahulu.....	15
BAB III METODELOGI PENELITIAN.....	21
3.1 Studi Literatur	22
3.2 Analisis Kebutuhan Pembangunan Model.....	22
3.2.1 Kebutuhan Data.....	22
3.2.2 Kebutuhan Harware dan Software	22
3.3 Model Klasifikasi	22
3.3.1 Pengumpulan Data.....	24
3.3.2 EDA dan Pra-proses Data.....	24
3.3.3 Pelabelan Data.....	24
3.3.4 Pra – proses Data.....	25
3.3.5 Pembagian Data Training dan Test.....	33
3.3.6 Pembobotan TF – IDF.....	33

3.3.7 Klasifikasi Naïve Bayes	34
3.3.8 Evaluasi Model Klasifikasi.....	34
BAB IV HASIL DAN PEMBAHASAN.....	35
4.1 Analisis Kebutuhan Pembangunan Model.....	35
4.1.1 Kebutuhan Data.....	35
4.1.1 Perangkat Keras (<i>Hardware</i>).....	35
4.1.2 Perangkat Lunak (<i>Software</i>).....	36
4.2 Pembangunan Model.....	36
4.2.1 Pengumpulan Data.....	36
4.2.2 EDA.....	38
4.2.3 Pelabelan Data.....	41
4.2.4 Pra – proses Data	42
4.2.5 Pembagian Data Uji dan Data Latih	59
4.2.6 Klasifikasi Data.....	60
4.2.7 Evaluasi Performa Model Klasifikasi.....	61
BAB V PENUTUP.....	62
5.1 Kesimpulan	65
5.2 Saran.....	65
DAFTAR PUSTAKA.....	66

DAFTAR GAMBAR

Gambar 3.1 <i>Flowchart</i> Metodologi Penelitian	21
Gambar 3.2 Diagram Alir Pembuatan Model	23
Gambar 3.3 Diagram Alir Pra – proses Data	26
Gambar 3.4 Diagram Alir Emotion Handling	30
Gambar 4.1 Contoh <i>Source Code</i> Pengumpulan Data	36
Gambar 4.2 <i>Box-plot</i> untuk kolom <i>Rate</i>	39
Gambar 4.3 Sebaran untuk kolom <i>Rate</i>	40
Gambar 4.4 <i>Source Code</i> Emoji Handling	43
Gambar 4.5 <i>Source Code</i> Emoticon Handling	44
Gambar 4.6 <i>Source Code</i> Case Folding	46
Gambar 4.7 <i>Source Code</i> Singkatan / Abbreviation Handling	47
Gambar 4.8 <i>Source Code</i> menghilangkan Tanda baca	49
Gambar 4.9 <i>Source Code</i> menghilangkan Angka.....	50
Gambar 4.10 <i>Source Code</i> Menghilangkan Huruf Tunggal.....	51
Gambar 4.11 <i>Source Code</i> Menghilangkan Huruf Berulang.....	53
Gambar 4.12 <i>Source Code</i> Tokenizing	54
Gambar 4.13 <i>Source Code</i> Stopword Removal	56
Gambar 4.14 <i>Source Code</i> Stemming.....	57
Gambar 4.15 <i>Source Code</i> Pembobotan TF - IDF.....	59
Gambar 4.16 <i>Source Code</i> Pembagian Data.....	59
Gambar 4.17 <i>Source Code</i> Klasifikasi Data Multinomial Naïve Bayes.....	60
Gambar 4.18 <i>Source Code</i> Klasifikasi Data Bernoulli Naïve Bayes.....	60
Gambar 4.19 <i>Source Code</i> Klasifikasi Data Gaussian Naïve Bayes	61

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	12
Tabel 2.2 Penelitian Terdahulu	17
Tabel 3.1 Daftar kata berdasarkan kelasnya	25
Tabel 3.2 Dictionary pada Emoticon Handling	27
Tabel 4.1 Contoh Ulasan yang didapatkan	37
Tabel 4.2 Ringkasan Statistic dari Ulasan Traveloka	38
Tabel 4.3 Menunjukkan adanya data <i>Missing Value</i> atau Tidak	40
Tabel 4.4 Contoh Ulasan yang sudah diberi Label	41
Tabel 4.5 Contoh Text yang melalui Emoji Handling	42
Tabel 4.6 Contoh Text Ulasan yang melalui Emoticon Handling	45
Tabel 4.7 Contoh Ulasan setelah proses Case Folding	46
Tabel 4.8 Contoh Ulasan setelah proses mengubah Singkatan	48
Tabel 4.9 Contoh Ulasan setelah proses menghilangkan Tanda Baca	49
Tabel 4.10 Contoh Ulasan setelah proses menghilangkan Angka	50
Tabel 4.11 Contoh Ulasan setelah proses menghilangkan Huruf Tunggal	52
Tabel 4.12 Contoh Ulasan setelah proses menghilangkan Huruf Berulang.....	53
Tabel 4.13 Contoh tweet setelah proses tokenizing	54
Tabel 4.14 Contoh tweet setelah proses stopword removal	56
Tabel 4.15 Contoh tweet setelah proses stemming	58
Tabel 4.16 Hasil Skenario	61
Tabel 4.17 Confusion Matrix	64