

**IMPLEMENTASI TEKNIK *WEB SCRAPING* UNTUK
KLASIFIKASI SENTIMEN MENGGUNAKAN
METODE *NAÏVE BAYES*
(STUDI KASUS DATA ULASAN JAKARTA
NOTEBOOK.COM *APP STORE*)**

SKRIPSI

Oleh :

RIO KHAIRUL IKHWAN

201710225026



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2021**

**IMPLEMENTASI TEKNIK *WEB SCRAPING* UNTUK
KLASIFIKASI SENTIMEN MENGGUNAKAN
METODE *NAÏVE BAYES*
(STUDI KASUS DATA ULASAN JAKARTA
NOTEBOOK.COM *APP STORE*)**

SKRIPSI

Oleh :

RIO KHAIRUL IKHWAN

201710225026



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2021**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Proposal Skripsi : Implementasi Teknik Web Scraping Untuk
Klasifikasi Sentimen Menggunakan Metode Naive
Bayes (Studi Kasus Data Ulasan Jakarta
Notebook.Com App Store)

Nama Mahasiswa : Rio Khairul Ikhwan

Nomor Pokok Mahasiswa : 201710225026

Program Studi/Fakultas : Informatika / Ilmu Komputer

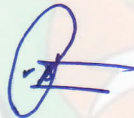
Tanggal Lulus Ujian Skripsi : 14 juli 2021

Bekasi, 19 Juli 2021

MENYETUJUI,

Pembimbing I

Pembimbing II



Ir. Muhammad Khaerudin, M.Kom
NIDN. 0413066604



Mukhlis, S.Kom., M.T
NIDN. 0312116802

LEMBAR PENGESAHAN

Judul Skripsi : Implementasi Teknik Web Scraping Untuk
Klasifikasi Sentimen Menggunakan Metode
Naïve Bayes (Studi Kasus Data Ulasan
Jakarta Notebook.com App Store)

Nama Mahasiswa : Rio Khairul ikhwan

Nomor Pokok Mahasiswa : 201710225026

Program Studi / Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 14 Juli 2021

Bekasi, 19 Juli 2021

Mengesahkan,

Ketua Tim Penguji : Joni Warta, S.Si., M.Si.
NIDN. 0317066202

Penguji (I) : Rasim, S.T., M.Kom.
NIDN. 0415027301

Penguji (II) : Ir. Muhammad Khaerudin, M.Kom.
NIDN. 0413066604

Mengetahui,

Ketua Program Studi
Informatika

Dekan
Fakultas Ilmu Komputer

Rakhmat Purnomo, S.Pd., S.Kom., M.Kom.
NIDN. 0322108201

Herlawati, S.Si., M.M., M.Kom.
NIDN. 0311097302



LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Rio Khairul Ikhwan
NPM : 201710225026
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Implementasi Teknik Web Scraping Untuk Klasifikasi Sentimen Menggunakan Metode Naive Bayes (Study Kasus Data Ulasan Jakarta Notebook.com App Store)

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 23 Juni 2021

Penulis



Rio Khairul Ikhwan

ABSTRAK

Rio Khairul Ikhwan. 201710225026. Implementasi Teknik *Web Scraping* Untuk Klasifikasi Sentimen Menggunakan Metode *Naïve Bayes* (Studi Kasus Data Ulasan Jakarta Notebook.com *App Store*)

App Store merupakan suatu layanan distribusi digital yang dioperasikan dan dikembangkan perusahaan *Apple*, layanan tersebut dapat diakses melalui website atau aplikasi yang tersedia pada *Ipod* dan *iPhone*. *App store* mempunyai sebuah fitur yang berisikan ulasan (*review*) keluhan atau saran dari para penggunanya. Ulasan dari pengguna aplikasi tersebut menjadi tolak ukur yang efektif dan efisien untuk menemukan informasi terhadap produk, yang dapat digunakan untuk melihat ulasan dari pengguna aplikasi. Ulasan atau *review* yang ada, terdapat banyaknya ulasan yang kurang tepat dengan rating yang diberikan oleh pengguna. Ulasan atau *review* yang diberikan pada umumnya berisikan keluhan-keluhan yang bersifat negatif yang akan mempengaruhi calon pengguna baru dalam mengunduh atau melakukan transaksi. Dalam penelitian ini, bertujuan untuk pengelompokan ulasan, keakuratan dan informasi yang didapat dari ulasan Jakarta notebook.com. Metode pengumpulan data yang berupa ulasan atau *review* yang diberikan pengguna kepada aplikasi Jakarta notebook.com dengan menggunakan teknik *scraping*. Pada proses klasifikasi sentimen yang dilakukan dengan menggunakan metode *Naïve Bayes*. Metode *Naïve Bayes* merupakan metode klasifikasi biner yang memanfaatkan probabilitas statistika yang sederhana dengan menerapkan aturan *Bayesian* menggunakan asumsi independen yang kuat. Hasil klasifikasi menggunakan model *Naïve Bayes Classifier* didapatkan tingkat akurasi sebesar 82% pada perbandingan 90% : 10%, akurasi sebesar 82.4% pada perbandingan 80% : 20%, dan akurasi sebesar 82.9% pada perbandingan 70% : 30%. Selanjutnya, pada proses asosiasi teks diperoleh informasi bahwa sentimen negatif pengguna aplikasi banyak membicarakan mengenai barang, rusak, murah, harga, dan beli. Kata tersebut biasanya menjadi tolak ukur untuk mengunduh aplikasi serta membeli produk. Hasil ulasan negatif tersebut dibuat dalam diagram *fishbone* untuk pemecahan masalah.

Kata kunci: Klasifikasi Sentimen, *Web Scraping*, *Naïve Bayes*, Jakarta notebook.

ABSTRACT

Rio Khairul Ikhwan. 201710225026. *Implementation of Web Scraping Technique for Sentiment Classification Using Naïve Bayes Method (Case Study Data Review Jakarta Notebook.com App Store)*

The App Store is a digital distribution service operated and developed by the Apple company, the service can be accessed through a website or application available on the Ipad and Iphone. The app store has a feature that contains reviews of complaints or suggestions from its users. Reviews from application users become an effective and efficient benchmark for finding information on products, which can be used to view reviews from application users. Existing reviews or reviews, there are many reviews that are not quite right with the rating given by the user. Reviews or reviews given generally contain negative complaints that will affect potential new users in downloading or making transactions. In this study, aims to group reviews, accuracy and information obtained from Jakarta notebook.com reviews. The method of collecting data is in the form of reviews or reviews given by users to the Jakarta notebook.com application using scraping techniques. The sentiment classification process is carried out using the Naïve Bayes method. The Naïve Bayes method is a binary classification method that utilizes simple statistical probabilities by applying Bayesian rules using strong independent assumptions. The results of the classification using the Naïve Bayes Classifier model obtained an accuracy rate of 82% at a comparison of 90%: 10%, an accuracy of 82.4% at a comparison of 80%: 20%, and an accuracy of 82.9% at a comparison of 70%: 30%. Furthermore, in the text association process, information is obtained that the negative sentiments of application users talk a lot about goods, damaged, cheap, price, and buying. The word is usually a benchmark for downloading applications and buying products. The results of these negative reviews are made in a fishbone diagram for problem solving.

keywords: *Sentiment Classification, Web Scraping, Naïve Bayes, Jakarta notebook.*

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK**

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Rio Khairul Ikhwan
NPM : 201710225026
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul :

“Implementasi Teknik Web Scraping Untuk Klasifikasi Sentimen Menggunakan Metode Naive Bayes (Studi Kasus Data Ulasan Jakarta Notebook.Com App Store)”

berserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Belaasi
Pada tanggal : 23 Juni 2021
Yang Menyatakan



(Rio Khairul Ikhwan)

KATA PENGANTAR

Segala Puji Bagi Allah SWT yang telah melimpahkan berkat dan rahmatnya, serta penulis hantarkan terimakasih kepada Baginda Nabi Muhammad SAW yang senantiasa memberikan syafaatnya kepada umatnya, sehingga penulis dapat melaksanakan penelitian dan menyelesaikan laporan kerja praktik ini.

Penulisan laporan ini penulis selesaikan berdasarkan data-data yang penulis amati selama pelaksanaan penelitian. laporan ini diajukan sebagai salah satu syarat untuk kelulusan Program Strata – 1 (S1) Informatika Universitas Bhayangkara Jakarta Raya.

Penulis menyadari dalam penulisan penelitian ini masih jauh dari kata sempurna, maka dari itu saran dan kritik sangat Penulis harapkan dan hargai dari berbagai pihak.

Dalam penelitian dan pengerjaan Skripsi/TA ini, penulis mendapatkan banyak bantuan dari berbagai pihak. Maka dari itu Penulis ingin mengucapkan terimakasih yang tidak terhingga kepada:

1. Bapak Dr. Bambang Karsono, S.H., M.M. Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Herlawati, S.Si.,M.Kom. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Rakhmat Poernomo, S.Pd., S.Kom., M.Kom. selaku Ketua Program Studi Informatika.
4. Bapak Ir.Muhammad Khaerudin,M.Kom. selaku dosen pembimbing Materi, yang telah memberikan bimbingan, masukan dan arahan dalam penyusunan Skripsi/Tugas akhir ini.
5. Bapak Mukhlis,S.Kom.,M.T. selaku dosen pembimbing Teknis dan Metodologi, yang telah membantu dan memberikan penulis arahan dalam penyusunan Skripsi/Tugas akhir ini.
6. Bapak Rakhmat Poernomo, S.Pd., S.Kom., M.Kom. selaku dosen pembimbing akademik.

7. Kedua orang tua, serta adik tercinta yang senantiasa selalu memberikan dukungan, doa, motivasi, serta kasih sayang maupun dukungan moril dan material kepada penulis.
8. Kepada sahabat-sahabat terkasih yang selalu menebarkan kebahagiaan serta memberi semangat untuk menyelesaikan Skripsi/TA ini .
9. Terimakasih teruntuk Lita apriliani yang berjuang dan selalu memotivasi saya untuk bersama-sama menempuh skripsi .
10. Terimakasih rekan dan sahabat sepejuangan, Rendy Dwi Sandy, Rido Nurahman Fauzan ,Muhammad Jefry Setiawan, Dwi Yoga Septia, Yoga Oktavianto serta seluruh teman kelas A1 , yang sudah membantu memberikan saran, masukan, motivasi semangat serta menjadi teman diskusi yang baik kepada penulis .
11. Semua pihak yang telah berkenan memberikan bantuan kepada penulis sehingga pembuatan laporan ini dapat diselesaikan dengan lancer.

Akhir kata, Penulis berharap Skripsi/TA ini dapat berguna di kemudian hari untuk berbagai pihak yang membutuhkannya.

Bekasi, ..27 Juni 2021

Penulis



(Rio Khairul Ikhwan)

DAFTAR ISI

	Halaman
LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Batasan Masalah.....	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	4
1.7 Tempat dan Waktu Penelitian	5
1.8 Metode Penelitian.....	5
1.8.1 Metode Analisis	5
1.8.2 Metode Perancangan	5
1.8.3 Metode Penujian.....	6
1.9 Sistematika penulisan	6
BAB II LANDASAN TEORI	8
2.1 Tinjauan Pustaka	8
2.2 Pengertian Model.....	10
2.3 App Store.....	10
2.4 Analisis Sentimen.....	10

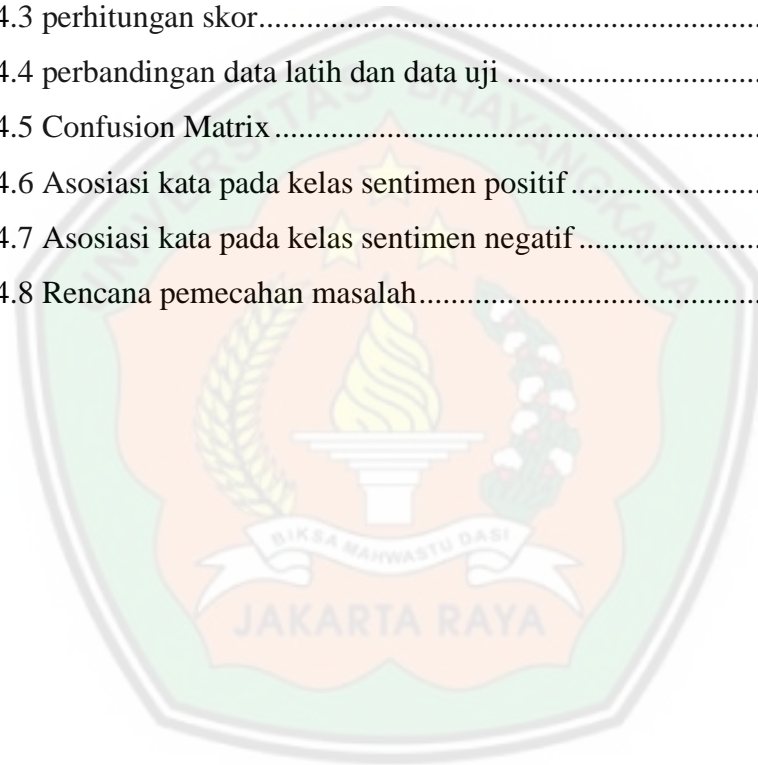
2.5	<i>Web Scraping</i>	11
2.6	Pengertian <i>Data Mining</i>	11
2.7	Pengertian Klasifikasi.....	12
2.8	Pengertian <i>Text Mining</i>	12
2.9	<i>Text Preprocessing</i>	12
2.10	Pengertian <i>Naïve Bayes</i>	13
2.11	Asosiasi Teks.....	13
2.12	Pengertian <i>Confusion Matrix</i>	14
2.13	Pengertian <i>Flowchart</i>	15
BAB III METODOLOGI PENELITIAN		19
3.1	Obyek Penelitian	19
3.2	Kerangka Penelitian.....	19
3.2.1	Identifikasi dan merumuskan masalah.....	20
3.2.2	Studi Pustaka.....	20
3.2.3	Pengumpulan data	20
3.2.4	<i>Text Preprocessing</i>	21
3.2.5	Pelabelan	22
3.2.6	Klasifikasi <i>Naïve Bayes</i>	22
3.2.7	Pengujian.....	22
3.3	Analisis Sistem berjalan	23
3.4	Sistem Usulan.....	24
3.5	Analisi Kebutuhan Sistem	25
3.5.1	Perangkat Keras	25
3.5.2	Perangkat Lunak.....	25
BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI		26
4.1	Hasil Penelitian.....	26
4.2	Pengumpulan Data.....	26
4.3	Analisis Deskriptif.....	30
4.4	<i>Text Preprocessing</i>	31
4.3.1	<i>Case Folding</i>	31
4.3.2	<i>Data Cleaning</i>	32
4.3.3	<i>Tokenizing</i>	33

4.3.4	<i>Spell normalization</i>	34
4.3.5	<i>Filtering</i>	34
4.5	Pelabelan Data	35
4.4.1	Skoring Data.....	36
4.4.2	Pelabelan Kelas Sentimen	36
4.6	Data Latih dan Data Uji.....	37
4.7	Klasifikasi dengan Metode <i>Naive Bayes</i>	39
4.8	Visualisasi.....	41
4.8.1	Ulasan Positif	41
4.8.2	Ulasan Negatif.....	43
4.9	Diagram <i>fishbone</i>	45
BAB V PENUTUP		48
5.1	Kesimpulan.....	48
5.2	Saran.....	49
DAFTAR PUSTAKA		50
LAMPIRAN		55



DAFTAR TABEL

	Halaman
Tabel 2. 1 Pemetaan Jurnal	8
Tabel 2.2 Model Confusion Matrix.....	14
Tabel 2.3 Simbol Flowchart Sistem.....	17
Tabel 2.4 Simbol Flowchart Program.....	18
Tabel 4.1 Data Review JakNote.....	29
Tabel 4.2 Contoh data Negatif dan Positif.....	35
Tabel 4.3 perhitungan skor.....	36
Tabel 4.4 perbandingan data latih dan data uji	38
Tabel 4.5 Confusion Matrix	40
Tabel 4.6 Asosiasi kata pada kelas sentimen positif.....	43
Tabel 4.7 Asosiasi kata pada kelas sentimen negatif.....	45
Tabel 4.8 Rencana pemecahan masalah.....	46



DAFTAR GAMBAR

	Halaman
Gambar 3.1 Flowchart Kerangka Penelitian	19
Gambar 3.2 Dataset Ulasan app store	21
Gambar 3.3 Flowchart analisis system berjalan.....	23
Gambar 3.4 Flowchart algoritma Naive Bayes	24
Gambar 4.1 Pembuatan kolom.....	26
Gambar 4.2 Pemilihan elemen nama	27
Gambar 4.3 Pemilihan elemen review	27
Gambar 4.4 Pemilihan elemen rating.....	28
Gambar 4.5 proses scraping	28
<i>Gambar 4.6 hasil scraping data ulasan</i>	<i>29</i>
Gambar 4.7 Grafik penilaian data ulasan Jakarta notebook.com.....	30
Gambar 4.8 hasil case folding.....	31
Gambar 4.9 hasil dari cleansing	32
Gambar 4.10 hasil penghapusan angka dan tanda baca	33
<i>Gambar 4.11 hasil dari tokenizing.....</i>	<i>33</i>
Gambar 4.12 hasil spell normalization.....	34
Gambar 4.13 hasil filtering	35
Gambar 4.14 pelabelan kelas sentimen.....	37
Gambar 4.15 Kata yang paling banyak muncul pada kelas positif	42
Gambar 4.16 wordcloud positif.....	42
Gambar 4.17 Kata yang paling banyak muncul pada kelas negatif	44
Gambar 4.18 wordcloud negatif.....	44
Gambar 4.19 Diagram fishbone komplain pengguna Jakarta notebook.com	46

DAFTAR LAMPIRAN

1. Script Pemrograman R
2. Stopword Bahasa Indonesia
3. Lembar Plagiarsm Checking
4. Biodata Mahasiswa
5. Kartu Bimbingan Skripsi Pembimbing 1
6. Kartu Bimbingan Skripsi Pembimbing 2

