



**ANALISIS SENTIMEN PADA ULASAN APLIKASI BUKUKAS DI
GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE
BAYES DENGAN SELEKSI FITUR CHI-SQUARE**

SKRIPSI

HARIYANTO PRASETYO

1810511058

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

INFORMATIKA

2022



**ANALISIS SENTIMEN PADA ULASAN APLIKASI BUKUKAS DI
GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE
BAYES DENGAN SELEKSI FITUR CHI-SQUARE**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

HARIYANTO PRASETYO

1810511058

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA**

FAKULTAS ILMU KOMPUTER

INFORMATIKA

2022

PERNYATAAN ORISINALITAS

Tugas Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun dirujuk saya nyatakan dengan benar.

Nama : Hariyanto Prasetyo

NIM : 1810511058

Tanggal : 26 Juli 2022

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 26 Juni 2022

Yang Menyatakan,



(Hariyanto Prasetyo)

PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Hariyanto Prasetyo

NIM : 1810511058

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non-Exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul:

Analisis Sentimen Pada Ulasan Aplikasi Bukukas Di Google Play Store Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-square

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti di Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih) media/formatkan, mengelola dalam bentuk pengkalan kata (Basis data), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 26 Juli 2022

Yang Menyatakan



(Hariyanto Prasetyo)

ANALISIS SENTIMEN PADA ULASAN APLIKASI BUKUKAS DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES DENGAN SELEKSI FITUR CHI-SQUARE

Hariyanto Prasetyo

ABSTRAK

Bukukas adalah sebuah aplikasi yang berfungsi untuk membantu menjadikan aktivitas mencatat semua transaksi, membuat laporan keuangan, dan mencatat menagih hutang piutang usaha dapat dilakukan secara daring. Aplikasi bukukas merupakan aplikasi yang fiturnya sangat bergantung pada pengguna yang memakainya, karena hal tersebut membuat kebanyakan keluhan para pengguna di *google play store* terhadap aplikasi bukukas adalah keluhan berupa fitur yang dihilangkan diminta untuk dikembalikan. Untuk mencari tahu permasalahan yang dialami oleh pengguna dan mempermudah dalam melakukan analisis untuk menentukan pengembangan aplikasi kedepannya maka akan dilakukan analisis sentimen dengan pendapat yang diperoleh dari *google play store*. Tahapan penelitian ini dimulai dengan *text processing* data dilanjutkan dengan *term weighting* menggunakan TF-IDF dan dilakukan proses *feature selection* menggunakan *chi square* dilanjutkan dengan klasifikasi kemudian hasil evaluasi. Hasil yang diperoleh dengan menggunakan metode *naïve bayes* dan *feature selection* menggunakan *chi square* diperoleh peningkatan akurasi sebesar 6% yaitu akurasi terbaik yang diperoleh adalah 81% jika menggunakan *feature selection*, untuk nilai *error rate*, *precision*, dan *recall* diperoleh sebesar 19% untuk *error rate*, 69% untuk nilai *precision*, dan 72% untuk nilai *recall*.

Kata Kunci: Bukukas, *naïve bayes*, analisis sentiment.

SENTIMENT ANALYSIS OF BUKUKAS APPLICATION REVIEWS ON GOOGLE PLAY STORE USING NAÏVE BAYES METHOD WITH CHI-SQUARE FEATURE SELECTION

Hariyanto Prasetyo

ABSTRACT

Bukukas is an application that serves to help make the activities of recording all transactions, making financial reports, and recording accounts payable can be done online. The Bukukas application is an application whose features are very dependent on the users who use it, because it makes most of the complaints from users on the Google Play Store against the Bukukas application are complaints in the form of features that are omitted being asked to be returned. To find out the problems experienced by users and make it easier to carry out an analysis to determine future application development, a sentiment analysis will be carried out with opinions obtained from the Google Play Store. The stages of this research begin with text processing data followed by term weighting using TF-IDF and feature selection process using chi square followed by classification and then evaluation results. The results obtained using the nave Bayes method and feature selection using chi square obtained an increase in accuracy of 6%, namely the best accuracy obtained is 81% when using feature selection, for the error rate, precision, and recall values obtained by 19% for the error rate, 69% for the precision value, and 72% for the recall value.

Keywords: Bukukas, nave Bayes, sentiment analysis.

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Hariyanto Prasetyo

NIM : 1810511058

Program Studi : S1 Informatika 2018

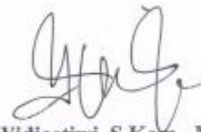
Judul : Analisis Sentimen Pada Ulasan Aplikasi Bukukas Di Google Play Store Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-square

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Didit Widiyanto, S.Kom., M.Si.

Penguji 1



Yuni Widiastwi, S.Kom., M.Si.

Penguji 2



Iin Emawati S.Kom., M.Si.

Pembimbing 1



Nurul Chamidah, S.Kom, M.Kom.

Pembimbing 2



Dr. Ermatita, M.Kom.

Dekan



Desta Sandva Prasvita, S.Kom., M.Kom.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal pengesahan :



KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT yang memberikan rahmat dan sehingga penulis mampu menyelesaikan skripsi berjudul “Analisis Sentimen Pada Ulasan Aplikasi Bukukas Di Google Play Store Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-square”. Penulis berterima kasih kepada :

1. Kedua orang tua penulis, Supriyadi (bapak) dan Kantun Sari Putri (ibu), serta seluruh keluarga sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Iin Emawati S.Kom., M.Si. dan Ibu Nurul Chamidah, S.Kom, M.Kom., selaku dosen pembimbing yang telah membimbing serta memberikan kritik dan saran sampai skripsi ini diselesaikan.
3. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
4. Bapak Desta Sandya Prasvita, S.Kom., M.Kom. selaku Ketua Prodi Informatika.
5. Bapak Henki Bayu Seta, S.Kom., MTI. selaku Dosen Pembimbing Akademik.
6. Seluruh dosen dan pegawai pada Fakultas Ilmu Komputer.
7. Teman-teman Informatika angkatan 2018.

Penulis juga mengucapkan terima kasih untuk seluruh pihak yang membantu penulisan skripsi ini.

Jakarta, 26 Juli 2022



(Hariyanto Prasetyo)

DAFTAR ISI

PERNYATAAN ORISINALITAS	i
PERNYATAAN PERSETUJUAN PUBLIKASI	ii
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR SIMBOL.....	xiii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	2
1.4. Luaran Penelitian.....	3
1.5. Ruang Lingkup	3
1.6. Manfaat Penelitian.....	4
1.7. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA	6
2.1. <i>Text Mining</i>	6
2.2. Analisis Sentimen.....	6
2.3. <i>Text Processing</i>	7

2.3.1.	<i>Noise Removal</i>	8
2.3.2.	<i>Case Folding</i>	8
2.3.3.	<i>Text Normalization</i>	8
2.3.4.	<i>Stemming</i>	9
2.3.5.	<i>Stopword Removal</i>	9
2.3.6.	<i>Tokenization</i>	9
2.4.	<i>Term Weighting</i>	10
2.4.1.	<i>Term Frequency-Invers Document Frequency (TF-IDF)</i>	10
2.5.	<i>Chi-square</i>	11
2.6.	<i>Naive Bayes Classifier</i>	12
2.7.	<i>Python</i>	13
2.8.	<i>Sastrawi Python</i>	13
2.9.	<i>Natural Language Toolkit</i>	14
2.10.	<i>Confusion Matrix</i>	14
2.11.	Review Riset yang Relevan	16
BAB 3 METODE PENELITIAN.....		18
3.1.	Kerangka Pikir.....	18
3.1.1.	Identifikasi Masalah	19
3.1.2.	Studi Literatur	19
3.1.3.	Pengumpulan Data	19
3.1.4.	Pelabelan Data.....	20
3.1.5.	<i>Text Preprocessing</i>	21
3.1.5.1.	<i>Noise Removal</i>	21
3.1.5.2.	<i>Case Folding</i>	22

3.1.5.3.	<i>Text Normalization</i>	22
3.1.5.4.	<i>Stemming</i>	22
3.1.5.5.	<i>Stopword Removal</i>	23
3.1.5.6.	<i>Tokenizing</i>	23
3.1.6.	Pembobotan Kata	24
3.1.7.	<i>Feature Selection</i>	24
3.1.8.	Klasifikasi	25
3.1.9.	Evaluasi	25
3.1.10.	Analisis Hasil Evaluasi.....	26
3.2.	Alat Bantu Penelitian.....	26
3.3.	Tempat Penelitian.....	27
3.4.	Jadwal Penelitian.....	27
BAB 4	HASIL DAN PEMBAHASAN	30
4.1	Pengumpulan Data	30
4.2	Pelabelan Data	31
4.3	<i>Text Preprocessing</i>	32
4.3.1	<i>Noise Removal</i>	33
4.3.2	<i>Case Folding</i>	34
4.3.3	<i>Text Normalization</i>	36
4.3.4	<i>Stemming</i>	38
4.3.5	<i>Stopword Removal</i>	39
4.3.6	<i>Tokenization</i>	41
4.4	<i>Term Weighting</i>	42
4.5	<i>Feature Selection (Chi-Squared)</i>	47

4.6	Klasifikasi.....	50
4.7	Evaluasi	56
4.8	Analisis Hasil Evaluasi.....	58
4.8.1	Hasil Performa Algoritma Menggunakan <i>Confusion Matrix</i>	59
4.8.2	Hasil Analisis Sentimen Bukukas	61
BAB 5 PENUTUP		68
5.1	Kesimpulan.....	68
5.2	Saran.....	69
DAFTAR PUSTAKA		70
RIWAYAT HIDUP.....		72
LAMPIRAN.....		73

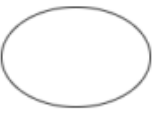


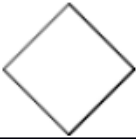

DAFTAR GAMBAR

Gambar 2.1 Proses Text Mining	6
Gambar 2.2 Proses Text Preprocessing	7
Gambar 2.3 Confusion Matrix	14
Gambar 3.1 Flowchart Penelitian	18
Gambar 3.2 Proses Klasifikasi	25
Gambar 4.1 Alamat Aplikasi Bukukas	30
Gambar 4.2 Data Komentar Pengguna Terhadap Aplikasi Bukukas	31
Gambar 4. 3 Flowchart Melakukan Noise Removal	33
Gambar 4.4 Flowchart Melakukan Case Folding	35
Gambar 4.5 Flowchart Melakukan Text Normalization	37
Gambar 4.6 Flowchart Melakukan Proses Stemming	38
Gambar 4.7 Flowchart Melakukan Proses Stopword Removal	40
Gambar 4.8 Flowchart Melakukan Proses Tokenization	41
Gambar 4.9 Hasil Melakukan Proses Term Weighting	47
Gambar 4.10 Hasil Chi Square Pendapat pengguna Bukukas	50
Gambar 4.11 Hasil Chi Square Taraf Nyata Terhadap Akurasi	56
Gambar 4.12 Hasil Chi Square Taraf Nyata Terhadap Fitur	57
Gambar 4. 13 Hasil Chi Square Fitur Terhadap Akurasi	58
Gambar 4.14 Hasil Persentase Sentimen Pendapat Pengguna	62
Gambar 4.15 Hasil Persentase Pendapat Pengguna Berdasarkan Bulan	62
Gambar 4.16 Persentase Sentimen Berdasarkan Versi Aplikasi	63
Gambar 4.17 Presentase Line Chart Sentimen Positif Pendapat Pengguna .	64
Gambar 4.18 Presentase Line Chart Sentimen Netral Pendapat Pengguna .	65
Gambar 4.19 Presentase Line Chart Sentimen Negatif Pendapat Pengguna	65
Gambar 4.20 Word Cloud Hasil Sentimen Kata Pendapat Pengguna	66

DAFTAR TABEL

Tabel 3.1 Contoh Data	20
Tabel 3.2 Pelabelan Data	20
Tabel 3.3 Proses Noise Removal	21
Tabel 3. 4 Proses Case Folding	22
Tabel 3.5 Proses Text Normalization	22
Tabel 3.6 Proses Stemming	23
Tabel 3.7 Proses Stopword Removal	23
Tabel 3.8 Proses Tokenizing	24
Tabel 3.9 Jadwal Penelitian	28
Tabel 4.1 Menentukan Pelabelan Pendapat Bukukas	32
Tabel 4.2 Hasil Proses Melakukan Noise Removal	33
Tabel 4.3 Hasil Proses Melakukan Case Folding	35
Tabel 4.4 Hasil Proses Melakukan Text Normalization	37
Tabel 4.5 Hasil Proses Melakukan Stemming	38
Tabel 4.6 Hasil Proses Melakukan Stopword Removal	40
Tabel 4.7 Hasil Proses Melakukan Tokenization	42
Tabel 4.8 Sampel Pendapat Pengguna Untuk Pembobotan Kata	43
Tabel 4.9 Perhitungan TF Pada Sampel Pendapat	43
Tabel 4.10 Perhitungan IDF pada dokumen	45
Tabel 4.11 Perhitungan TF-IDF pada dokumen	46
Tabel 4.12 Proses Perhitungan Chi Square Menggunakan Kata Catat	47
Tabel 4. 13 Hasil Perhitungan Chi Square Dengan Data Tabel 4.8	48
Tabel 4. 14 Pembagian Data Latih Dan Data Uji	50
Tabel 4.15 Nilai Peluang Tiap Kata Untuk Setiap Kelas	52
Tabel 4.16 Sampel Data Uji	53
Tabel 4.17 Perhitungan IDF Pada Data Uji	54
Tabel 4.18 Tabel Confusion Matrix	59

DAFTAR SIMBOL

Simbol Flowchart			
NO	GAMBAR	NAMA	KETERANGAN
1		<i>Terminal (start, end)</i>	Menggambarkan bagaimana kegiatan dimulai atau kegiatan berakhir.
2		<i>Flow Direction</i>	Menggambarkan hubungan antar simbol yang menyatakan suatu jalannya proses dalam sistem.
3		<i>Process</i>	Menggambarkan deskripsi dari proses yang dijalankan.
4		<i>Decision</i>	Menggambarkan proses yang memiliki kondisi dimana harus memilih lebih dari satu proses.
5		<i>Data</i>	Menggambarkan proses yang dilakukan yang memiliki data masukan (<i>input</i>) atau keluaran (<i>output</i>).