



**ANALISIS SENTIMEN PADA *REVIEW* PENGGUNA *E-COMMERCE* BIDANG PANGAN MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE
(Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)**

SKRIPSI

GEYESSELLA MANIK

1710511004

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021



**ANALISIS SENTIMEN PADA *REVIEW* PENGGUNA *E-COMMERCE* BIDANG PANGAN MENGGUNAKAN METODE
*SUPPORT VECTOR MACHINE***

(Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)

SKRIPSI

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

GEYESSELLA MANIK

1710511004

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021

PERNYATAAN ORISINALITAS

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

Nama : Geyessella Manik

NIM : 1710511004

Tanggal : 20 Juli 2021

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

Jakarta, 20 Juli 2021

Yang Menyatakan,



(Geyessella Manik)

PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta. Saya yang bertanda tangan dibawah ini :

Nama : Geyessella Manik

NIM : 1710511004

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui 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 *Review* Pengguna *E-Commerce* Bidang Pangan Menggunakan Metode *Support Vector Machine* (Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan kata (Basis data), merawat dan memublikasikan 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 : 20 Juli 2021

Yang menyatakan,



(Geyessella Manik)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut :

Nama : Geyessella Manik
NIM : 1710511004
Program Studi : Informatika
Judul : Analisis Sentimen Pada *Review* Pengguna *E-Commerce*
Bidang Pangan Menggunakan Metode *Support Vector Machine*
(Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)

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.



Yuni Widiastiwi, S.Kom., M.Si.

Penguji I



Iin Ernawati, S.Kom., M.Si.

Pembimbing I



Dr. Ernawati, M.Kom.

Dekan



Nurul Chamidah, S.Kom., M.Kom.

Penguji II



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.

Pembimbing II



Yuni Widiastiwi, S.Kom., M.Si.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 20 Juli 2021



**ANALISIS SENTIMEN PADA *REVIEW* PENGGUNA *E-COMMERCE*
BIDANG PANGAN MENGGUNAKAN METODE *SUPPORT VECTOR
MACHINE***

(Studi Kasus: *Review* Sayurbox dan Tanihub pada Google Play)

GEYESSELLA MANIK

ABSTRAK

Sayurbox dan Tanihub merupakan *startup digital* yang membantu masyarakat dalam memenuhi kebutuhan dapur dengan menjual hasil pertanian secara online. Citra dari opini publik berupa kesan positif dan negatif merupakan hal yang penting yang biasanya ingin diketahui sebuah perusahaan karena dapat mempengaruhi daya tarik perusahaan serta berpengaruh terhadap efektifitas pemasaran produk atau layanan perusahaan. Opini dari publik dimuat dengan jumlah yang banyak, sehingga diperlukan suatu teknik yang dapat mengelompokkan opini tersebut secara otomatis dengan cepat. Teknik yang dapat digunakan pada bisnis untuk menentukan opini publik termasuk positif atau negatif adalah aplikasi *text mining*. Salah satu bidang *text mining* yang biasa digunakan yaitu analisis sentimen yang berguna untuk mendapatkan suatu informasi dengan memantau opini pengguna terhadap produk atau jasa yang diberikan perusahaan tersebut. Sayurbox dan Tanihub merupakan dua *e-commerce* yang berbeda namun menyediakan layanan dengan bidang yang sama yang dapat diakses dengan mengunduhnya dari situs *Google Play*. Dalam melakukan proses analisis sentimen, data mentah diperoleh dengan melakukan penarikan data *rating* dan *review* dari situs *google play*. Selanjutnya data dibersihkan dari kata dan simbol yang tidak relevan dengan sentimen lalu melabelinya kedalam dua kelas yakni positif dan negatif. Sebelum masuk ke proses klasifikasi, data review yang masih bernilai kualitatif diubah dahulu menjadi data yang bernilai kuantitatif dengan penggunaan metode *Term Frequency-Invers Document Frequency*(TF IDF). Hasil dari pengklasifikasian dengan metode *Support Vector Machine*(SVM) dengan jangka waktu Februari 2020

hingga Januari 2021 diperoleh besaran akurasi tertinggi yaitu pada data *review* sayurbox sebesar 91,4% dengan jumlah sentimen terbanyak adalah sentimen positif sebanyak 738 (70%). Sedangkan pada data *review* Tanihub, total akurasi sebesar 88,8% dengan sentimen terbanyak yaitu sentimen positif berjumlah 348 (65%). Adapun proporsi pembagian data yang digunakan adalah 80 : 20.

Kata Kunci : Analisis Sentimen, *E-commerce*, Sayurbox, Tanihub, *Text Mining*, *SVM Classifier*, *Google Play*.

**SENTIMENT ANALYSIS OF USER REVIEWS OF E-COMMERCE IN
THE FOOD SECTOR USING SUPPORT VECTOR MACHINE
(Case Study: Review of Sayurbox and Tanihub on Google Play)**

GEYESSELLA MANIK

Abstract

Sayurbox and Tanihub are digital startups that help people by selling agricultural products online. The image of public opinion in the form of positive and negative impressions is an important thing that a company usually wants to know because it can affect the attractiveness of the company and affect the effectiveness of marketing the company's products or services. Opinions from the public are loaded in large numbers, so we need a technique that can group those opinions automatically quickly. Techniques that can be used in businesses to determine public opinion, including positive or negative are text mining applications. One of the fields of text mining that is commonly used is sentiment analysis which is useful for obtaining information by monitoring user opinions on the products or services provided by the company. Sayurbox and Tanihub are two different e-commerce sites but provide services in the same field that can be accessed by downloading them from the Google Play site. In conducting the sentiment analysis process, raw data is obtained by withdrawing rating and review data from the Google Play site. Furthermore, the data is cleaned of words and symbols that are not relevant to the sentiment and then labeled it into two classes, namely positive and negative. Before entering the classification process, the review data which is still of qualitative value is converted into data of quantitative value by using the Term Frequency Inverse Document Frequency (TF IDF method). The results of the classification using the Support Vector Machine (SVM method) with a period of February 2020 to January 2021 obtained the highest accuracy is in the data review sayurbox of 91.4% with the highest number of sentiments being positive sentiment as much as 738 (70%).

While in the data review tanihub, the total accuracy is 88.8% with the most positive sentiment, which is 348 (65%). The proportion of data sharing used is 80 : 20.

Keywords : Sentiment Analysis, E-commerce, Sayurbox, Tanihub, Text Mining, SVM Classifier, Google Play.

KATA PENGANTAR

Puji syukur yang sebesar-besarnya penulis panjatkan kepada Tuhan Yang Maha Kuasa yang selalu menjadi penopang dan teman terbaik penulis. Dengan anugerah-Nya, akhirnya penulis dapat menyelesaikan skripsi ini yang berjudul “Analisis Sentimen Pada *Review* Pengguna *E-Commerce* Bidang Pangan Menggunakan Metode *Support Vector Machine* (Studi Kasus: *Review* Sayurbox Dan Tanihub Pada Google Play)”. Dan tak lupa juga, penulis ingin ucapkan terima kasih kepada :

1. Kedua Orang tua, beserta keluarga yang senantiasa ada memberi dukungan dan nasehat.
2. Ibu Iin Ernawati, S.Kom., M.Si. selaku Dosen Pembimbing yang telah memberikan saran dan bimbingan yang sangat bermanfaat dari awal pembuatan Proposal hingga penyelesaian Skripsi.
3. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. selaku Dosen Pembimbing yang telah memberikan saran dan bimbingan yang sangat bermanfaat dari awal pembuatan Proposal hingga Skripsi.
4. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
5. Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku Ketua Prodi Informatika.
6. Bapak Bayu Hananto, S.Kom., M.Kom. selaku Dosen Pembimbing Akademik, dan juga seluruh Bapak, Ibu Dosen Pengajar Fakultas Ilmu Komputer UPN Veteran Jakarta.
7. Teman-teman seperjuangan Informatika, FIK 2017 sekaligus teman terdekat yaitu Azini Putri dan Farhana Nabila.
8. Teman-teman KSM Android UPN Veteran Jakarta, khususnya Periode 2019/2020 dan 2020/2021.

Dan seluruh pihak yang banyak membantu penulis dalam menyelesaikan Skripsi ini.

Jakarta, 20 Juli 2021

DAFTAR ISI

HALAMAN AWAL	i
JUDUL	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT	viii
KATA PENGANTAR	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Ruang Lingkup	3
1.4. Tujuan Penelitian	4
1.5. Luaran yang Diharapkan	4
1.6. Manfaat	4
3.1. Sistematika Penulisan	5
BAB 2 TINJAUAN PUSTAKA	6
2.1. <i>Electronic Commerce</i>	6
2.2. Sayurbox dan Tanihub	7
2.3. <i>Text Mining</i>	7
2.4. Aplikasi <i>Text Mining</i>	8

2.4.1.	Analisis Sentimen	8
2.5.	<i>Text Preprocessing</i>	9
2.5.1.	<i>Translation</i> (Penerjemahan kata dan simbol)	9
	Kamus Emoticon.....	10
2.5.2.	<i>Case Folding</i>	10
2.5.3.	<i>Spelling Normalization</i>	10
	Kamus Slangword	10
2.5.4.	<i>Tokenizing</i>	11
2.5.5.	<i>Filtering</i>	11
	Kamus Stopword.....	11
2.6.	Ekstraksi TF IDF.....	11
2.7.	<i>Machine Learning</i>	13
2.8.	Klasifikasi SVM.....	13
2.9.	Ketepatan Klasifikasi	14
2.10.	Visualisasi	15
2.11.	Review Penelitian Terdahulu	16
2.12.	Kesimpulan peneliti terdahulu	18
BAB 3 METODOLOGI PENELITIAN.....		21
3.1	Kerangka Pikir	21
3.1.1.	Menentukan Topik	21
3.1.2.	Identifikasi dan Perumusan Masalah	22
3.1.3.	Tujuan	22
3.1.4.	Studi Literatur	22
3.1.5.	<i>Scraping Data</i>	22
3.1.6.	Praproses teks.....	23
3.1.7.	Pelabelan kelas sentiment	23

3.1.8.	Membagi Data.....	24
3.1.9.	Ekstraksi Fitur	24
3.1.10.	Proses Klasifikasi.....	24
3.1.11.	Hasil	24
3.1.12.	Laporan hasil akhir.....	25
3.2.	Alat Pendukung.....	25
3.3.	Waktu dan Tempat	25
3.4.	Jadwal	25
BAB 4 HASIL DAN PEMBAHASAN.....		27
4.1.	Data	27
4.1.1.	Hasil scraping.....	30
4.2.	Praproses data teks	33
4.2.1.	<i>Translation</i>	33
4.2.2.	<i>Case Folding</i>	35
4.2.3.	<i>Spelling Normalization</i>	36
4.2.4.	<i>Tokenizing</i>	38
4.2.5.	<i>Filtering</i>	39
4.3.	Pelabelan kelas sentiment	41
4.4.	Bagi data	44
4.5.	Simulasi perhitungan TF IDF	45
4.6.	Proses Klasifikasi.....	47
4.7.	Hasil Klasifikasi Sentiment pada bulan April dan Mei	49
4.8.	Visualisasi.....	51
BAB 5 PENUTUP		56
5.1.	Kesimpulan	56
5.2.	Saran	57

DAFTAR PUSTAKA	58
RIWAYAT HIDUP	1
LAMPIRAN	2

DAFTAR GAMBAR

Gambar 1. 1 Tiga dampak akibat coronavirus	1
Gambar 2. 1 Proses umum <i>Text Mining</i>	8
Gambar 2. 2 Tahap Praproses Teks	9
Gambar 2. 3 SVM pada kelas yang linear	14
Gambar 2. 4 SVM pada kelas yang tidak linear	14
Gambar 2. 5 Contoh visualisasi berbentuk wordcloud	16
Gambar 3. 1 Diagram Tahapan Penelitian	21
Gambar 4. 1 Modul <i>python</i> untuk <i>scraping</i>	27
Gambar 4. 2 <i>Website</i> Sayurbox – Klik Panen Kirim	28
Gambar 4. 3 <i>Website</i> Tanihub – Belanja & Dukung Petani.....	28
Gambar 4. 4 kode <i>scraping</i> data	29
Gambar 4. 5 Data mentah.....	30
Gambar 4. 6 Grafik ulasan pada Sayurbox	31
Gambar 4. 7 Grafik ulasan pada Tanihub	31
Gambar 4. 8 Persentase rating Sayurbox	32
Gambar 4. 9 Persentase rating Tanihub	32
Gambar 4. 10 <i>Flowchart translation</i> ulasan	34
Gambar 4. 11 <i>Flowchart</i> proses <i>case folding</i>	35
Gambar 4. 12 <i>Flowchart</i> proses normalisasi teks	37
Gambar 4. 13 <i>Flowchart</i> proses tokenisasi kalimat	38
Gambar 4. 14 Sebelum dan sesudah tokenisasi	39
Gambar 4. 15 <i>Flowchart</i> proses <i>filtering</i>	39
Gambar 4. 16 Jumlah Kelas Sentiment pada Sayurbox	43
Gambar 4. 17 Jumlah Kelas Sentiment pada Tanihub	43
Gambar 4. 18 Hasil analisa Dukungan dan Keluhan pada April menggunakan visualisasi <i>Bar Chart</i>	49
Gambar 4. 19 Hasil analisa Dukungan dan Keluhan pada Mei menggunakan visualisasi <i>Bar Chart</i>	50
Gambar 4. 20 Hasil analisa Dukungan dan Keluhan pada April menggunakan visualisasi <i>Pie Chart</i>	51

Gambar 4. 21 Hasil analisa Dukungan dan Keluhan pada Mei menggunakan visualisasi <i>Pie Chart</i>	51
Gambar 4. 22 <i>WordCloud</i> pada <i>review</i> Sayurbox.....	52
Gambar 4. 23 Frekuensi kata yang paling banyak muncul pada <i>review</i> Sayurbox53	
Gambar 4. 24 <i>WordCloud</i> pada <i>review</i> Tanihub.....	54
Gambar 4. 25 Frekuensi kata yang paling banyak muncul pada <i>review</i> Tanihub	55

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	15
Tabel 2. 2 Kesimpulan Penelitian terdahulu	18
Tabel 3. 1 Jadwal Kegiatan Penelitian.....	25
Tabel 4. 1 Kelas pada <i>website</i> yang digunakan untuk <i>scraping</i>	29
Tabel 4. 2 Sebelum dan sesudah penerjemahan.....	34
Tabel 4. 3 Penyeragaman huruf	35
Tabel 4. 4 Normalisasi teks.....	37
Tabel 4. 5 Hasil sebelum dan sesudah tahap filtrasi	40
Tabel 4. 6 Contoh <i>Inconsistent rating</i> terhadap <i>review</i>	42
Tabel 4. 7 pembagian data <i>review</i> sayurbox	44
Tabel 4. 8 Pembagian data <i>review</i> tanihub.....	44
Tabel 4. 9 Sampel 3 Dokumen sebelum dan sesudah pembersihan data.....	45
Tabel 4. 10 Hasil perhitungan TF IDF tiga sampel.....	46
Tabel 4. 11 Nilai akurasi <i>machine learning</i> dengan metode SVM.....	47
Tabel 4. 12 <i>Confusion matrix</i> klasifikasi SVM pada Sayurbox data.....	48
Tabel 4. 13 <i>Confusion matrix</i> klasifikasi SVM pada Tanihub data.....	48