



**ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE DENGAN
KOMENTAR FOLLOWERS MENGGUNAKAN *TOOLS ORANGE DATA MINING***

SKRIPSI

**AYUNI SENTIYA
151410311**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
PALEMBANG
2019**



**ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE DENGAN
KOMENTAR FOLLOWERS MENGGUNAKAN *TOOLS ORANGE DATA MINING***

**AYUNI SENTIYA
151410311**

**Skripsi ini diajukan sebagai syarat memperoleh gelar
Sarjana Komputer**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINA DARMA
PALEMBANG
2019**

HALAMAN PENGESAHAN

ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE DENGAN KOMENTAR FOLLOWERS MENGGUNAKAN TOOLS ORANGE DATA MINING

**AYUNI SENTIYA
151410311**

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Pembimbing

Heri Suroyo, S.Si., M.Kom.

Palembang, 30 Juli 2019
Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Bina Darma

Ketua,

Dedy Syamsuar, Ph.D.



HALAMAN PERSETUJUAN

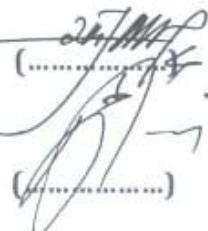
Skripsi Berjudul "*Analisis Text Clustering Akun Fanpage Shopee Dengan Komentar Followers Menggunakan Tools Orange Data Mining*" Oleh "Ayuni Sentiya", telah dipertahankan didepan komisi penguji pada Hari Selasa Tanggal 30 Bulan Juli Tahun 2019.

Komisi Penguji

1. Ketua : Heri Suroyo, M.Kom.



2. Anggota : Dr. Tri Basuki K, S.Kom., M.Eng



3. Anggota : Andri, M.Cs.



Mengetahui,
Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Bina Darma

Ketua,



Dr. Edi Surya Negara, M.Kom

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Ayuni Sentiya

NIM : 151410311

Dengan ini menyatakan bahwa :

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar sarjana di Universitas Bina Darma atau diperguruan tinggi lain;
2. Skripsi ini murni gagasan, rumusan, dan penelitian saya sendiri dengan arahan Tim Pembimbing;
3. Didalam Skripsi ini tidak terdapat karya dan pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang memasukkan ke dalam daftar rujukan;
4. Saya bersedia skripsi yang saya hasilkan ini di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses publik secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, Juli 2019
Yang membuat pernyataan,



MOTTO DAN PERSEMPAHAN

MOTTO:

- ✓ **Hidup itu bukan seperti mengetik pada *computer*, yang mana anda bisa menghapusnya sesuka hati dengan tombol *backspace*. Namun, anda harus memikirkan jalan hidup anda dengan merencanakan sesuatu yang baik dan terstruktur.**
- ✓ **Dunia ini bersifat global dan anda bersifat individual, tentu saja tidak akan pernah terhubung. Sebaiknya ubahlah pola pikir, maka anda dapat mengubah dunia.**
- ✓ **Tidak akan ada sebuah hasil yang sia-sia nantinya karena ilmu yang anda sedang pelajari saat ini akan tumbuh meraih kesuksesanmu.**

KUPERSEMBAHKAN KEPADA:

- ☞ Orang Tua saya, Ibu Resmala Dewi dan Bapak Zulheri yang selalu memberikan do'a dan mendukung saya hingga di titik ini.
- ☞ Adik laki-lakiku satu-satunya saudaraku, Julian Adji Satria yang telah menjadi pelipur lara selama saya menempuh pendidikan hingga di titik ini.
- ☞ Bapak Heri Suroyo, M.Kom. selaku pembimbing yang telah membimbing saya dalam penyusunan penulisan skripsi ini.
- ☞ Keluarga besarku yang telah memberikan nasehat, do'a, dan semangat.
- ☞ Sahabatku Nurul Fauziah, Dea Rizky Saputri, Citra Putri, Fitri Nur Aqidah, dan Dewi Zelika yang telah menjadi rumah kedua saya yang selalu ada saat suka maupun duka.
- ☞ Teman-teman seperjuangan, Rany Veranti, Fadhila Rizky Fiki Amelia, Ikrimah, Febby Anggraini, dan Dea Apriani tempat berkeluh kesah dan tempat melimpahkan kebahagiaan selama saya menempuh pendidikan S1 ini.
- ☞ Semua teman seperjuangan angkatan 2015 Sistem Informasi.
- ☞ Seluruh Dosen dan Karyawan Universitas Bina Darma Palembang.

☞ Segenap lingkungan Universitas Bina Darma Palembang yang memberikan fasilitas dan prasarana yang baik untuk dapat menyelesaikan skripsi ini.

ABSTRAK

Hasil penelitian ini ditemukan fakta bahwa *followers* akun *twitter* Shopee Indonesia paling banyak merespon status yang berhubungan dengan kata kuis berhadiah, *retweet* yang memberikan *handphone*, dan tebakan dalam bentuk *vote* dimana topik pembicaraan dalam sebuah *tweet* yang beragam membuat perhatian untuk diteliti. *Text mining* yang digunakan untuk ekstrak secara otomatis dari sumber data teks yang berbeda yang menangani masalah *clustering*. Dengan menggunakan metode *K-Means Clustering* sebagai metode pengelompokan data yang mempartisi data ke dalam bentuk dua atau lebih kelompok. Penelitian ini mencoba menggunakan *text mining* dalam proses untuk pengelolaan serta peringkasan teks sehingga menghasilkan analisis *text clustering* terhadap akun *fanpage* Shopee Indonesia dengan menggunakan *tools Orange Data Mining* dengan memanfaatkan *Preprocess Text* yang meliputi *transformation*, *tokenization*, *normalization*, dan *filtering* yang bertujuan agar teks dapat dibaca dan dianalisis dari *Orange Data Mining* sehingga menghasilkan topik pembahasan dari kata yang dominan muncul dari status dan komentar *followers* Shopee Indonesia. Kumpulan teks dalam jumlah banyak akan di *scraping* dengan menggunakan *Web Scraper*. Dari penelitian yang dilakukan mendapatkan kesimpulan bahwa Shopee Indonesia harus lebih membahas tentang kuis berhadiah, *retweet* yang memberikan *handphone*, dan tebak dalam bentuk *vote*.

Kata Kunci : Twitter, Analisis Text Clustering, Orange Data Mining, Web Scraper, Shopee Indonesia.

ABSTRACT

The results of this study found the fact that Shopee Indonesia Twitter account followers most responded to the status associated with the word prize quiz, retweet that gave a cellphone, and guesses in the form of votes where the topic of conversation in a diverse tweet made attention to be researched. Text mining is used to extract automatically from different text data sources that handle clustering problems. By using the K-Means Clustering method as a method of grouping data that partition data into two or more groups. This study tries to use text mining in the process of management and text summarization so as to produce analysis of text clustering on Shopee Indonesia fanpage accounts using Orange Data Mining tools by utilizing Preprocess Text which includes transformation, tokenization, normalization, and filtering that aims to make text readable and analyzed from Orange Data Mining so as to produce a discussion topic of the dominant words emerging from the status and comments of Shopee Indonesia followers. A large number of texts will be scraped by using a Web Scraper. From the research conducted, it was concluded that Indonesian Shopee should discuss more about prize quizzes, retweets that give mobile phones, and guess in the form of votes.

Keywords: Twitter, Text Clustering Analysis, Orange Data Mining, Web Scraper, Shopee Indonesia.

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya skripsi ini dapat diselesaikan dengan baik untuk memenuhi salah satu syarat mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Bina Darma. Dalam penulisan skripsi ini, tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasnya pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi penelitian ini, terutama kepada:

1. Ibu Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang.
2. Bapak Dedy Syamsuar, Ph.D. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Edi Surya Negara, M.Kom. selaku Ketua Program Studi Sistem Informasi.
4. Bapak Heri Suroyo, M.Kom. selaku Pembimbing yang telah memberikan bimbingan dalam penulisan skripsi ini.
5. Kedua Orang Tua saya, saudara laki-laki saya satu-satunya, keluarga besar, sahabat, serta teman-teman yang telah memberikan semangat.
6. Segenap lingkungan Universitas Bina Darma Palembang.

Palembang, Juli 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
 BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan dan Manfaat.....	4
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Metodologi Penelitian.....	6
1.5.1 Metode Penelitian.....	6
1.5.2 Metode Pengumpulan Data.....	7
1.5.3 Metode Pengolahan Data.....	8
1.5.4 Metode Analisis Data.....	11
1.6 Sistematika Penulisan.....	12
 BAB II TINJAUAN PUSTAKA	
2.1 Tinjauan Umum.....	14
2.1.1 Profil Pengguna Shopee	14
2.2 Landasan Teori.....	16
2.2.1 <i>Data Mining</i>	17
2.2.1.1 Tahapan <i>Data Mining</i>	18
2.2.1.2 Pengelompokkan <i>Data Mining</i>	21
2.2.2 <i>Text Mining</i>	22
2.2.3 <i>Clustering</i>	23
2.2.4 <i>K-Means</i>	24
2.2.5 <i>Orange Data Mining</i>	25
2.2.6 <i>Web Scraping</i>	26

BAB III ANALISIS DAN RANCANGAN

3.1 Tempat dan Waktu Penelitian.....	29
3.2 Alat dan Bahan.....	29
3.3 Data Penelitian	30
3.3.1 Sumber Data.....	31
3.3.2 Pengumpulan Data.....	31
3.3.3 Rancangan <i>Data Mining</i>	37
3.3.4 Rancangan Proses <i>K-Means Clustering</i>	42
3.3.5 Analisis <i>Clustering</i>	43
3.3.5.1 <i>Box Plot Clustering</i>	43
3.3.5.2 <i>Scatter Plot Clustering</i>	44

BAB IV HASIL DAN PEMBAHASAN

4.1 <i>Design Widget Text Clustering</i>	45
4.2 Data Inputan (<i>Corpus</i>)	46
4.3 <i>Corpus Viewer</i>	46
4.4 <i>Preprocessing Text</i>	47
4.4.1 Hasil <i>Preprocess Text</i>	51
4.5 Hasil Analisis <i>Text Clustering</i>	55

BAB V PENUTUP

5.1. Kesimpulan	59
5.2. Saran	59

LAMPIRAN

DAFTAR GAMBAR

	Halaman
Gambar 2.1 <i>E-commerce</i> Yang Paling Banyak Dicari Di <i>Google Trends</i>	15
Gambar 2.2 Tahapan <i>KDD (Knowledge Discovery in Database)</i>	18
Gambar 2.3 Proses Kerja Aplikasi Pengolah Data.....	26
Gambar 2.4 Ilustrasi Cara Kerja <i>Web Scraper</i>	27
Gambar 3.1 Ekstensi <i>Web Scraper</i> ke <i>Chrome</i>	32
Gambar 3.2 Tampilan <i>Web Scraper</i>	32
Gambar 3.3 Tampilan <i>Import Sitemap</i>	33
Gambar 3.4 Tampilan Kode <i>Sitemap JSON</i>	33
Gambar 3.5 Tampilan Data Yang Telah Di <i>Import Sitemap</i>	34
Gambar 3.6 Tampilan <i>Advanced Search Data Tweet Historis</i>	34
Gambar 3.7 Tampilan <i>Link</i> Yang Akan Di <i>Scraping</i>	35
Gambar 3.8 <i>Scraping</i> Status dan Komentar Menggunakan <i>Web Scraper</i>	36
Gambar 3.9 Tampilan Data Yang Ingin Di <i>Scraping</i>	36
Gambar 3.10 Perintah <i>Download File</i>	37
Gambar 3.11 Data Hasil <i>Scraping</i> Dari <i>Twitter Shopee Indonesia</i>	38
Gambar 3.12 <i>Data Cleaning</i> Akun <i>Twitter Shopee Indonesia</i>	39
Gambar 3.13 Proses <i>Data Selection Status @ShopeeID</i>	40
Gambar 3.14 <i>Design Widget K-Means Clustering</i>	42
Gambar 3.15 Hasil <i>Clustering K-Means</i>	43
Gambar 3.16 Tampilan <i>Box Plot Clustering</i>	44
Gambar 3.17 Tampilan Hasil <i>Scatter Plot Clustering</i>	44
Gambar 4.1 <i>Design Widget Text Clustering</i>	45
Gambar 4.2 <i>Preprocess Text</i> Pada <i>Orange Data Mining</i>	47
Gambar 4.3 Hasil <i>Word Cloud Cluster 1</i> Status <i>Shopee</i>	51
Gambar 4.4 <i>Most Frequent Word Cluster 1</i> Status <i>Shopee</i>	52
Gambar 4.5 Hasil <i>Word Cloud Cluster 2</i> Status <i>Shopee</i>	52
Gambar 4.6 <i>Most Frequent Word Cluster 2</i> Status <i>Shopee</i>	53
Gambar 4.7 Hasil <i>Word Cloud Cluster 3</i> Status <i>Shopee</i>	53
Gambar 4.8 <i>Most Frequent Word Cluster 3</i> Status <i>Shopee</i>	54
Gambar 4.9 Hasil <i>Word Cloud Komentar Followers Shopee</i>	54
Gambar 4.10 <i>Most Frequent Word Komentar Shopee</i>	55

DAFTAR TABEL

	Halaman
Tabel 2.1 Jumlah Pengikut <i>E-Commerce</i> Di Media Sosial	14
Tabel 3.1 <i>Link</i> Akun Media Sosial.....	31
Tabel 3.2 Nama Atribut/ <i>Field</i>	41
Tabel 4.1 Daftar <i>Stopwords</i> Dalam Setiap <i>Cluster</i> Status Shopee.....	50
Tabel 4.2 Daftar <i>Stopwords</i> Komentar <i>Followers</i>	50
Tabel 4.3 Hasil Pengelompokkan Data Konten Status Shopee	56
Tabel 4.4 Hasil Pengelompokkan Data Komentar <i>Followers</i> Shopee	57