

# **Klasterisasi *Online Shop* Berdasarkan *Caption* dengan Algoritma**

## ***Jaro Winkler Distance* dan *K-Nearest Neighbor***

Diajukan Sebagai Syarat Untuk Menyelesaikan

Pendidikan Program Strata-1 Pada

Jurusan Teknik Informatika



Oleh:

Hikmah Jusadi

09021181320059

JURUSAN TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2021

## LEMBAR PENGESAHAN TUGAS AKHIR

### Klasterisasi *Online Shop* Berdasarkan *Caption* dengan Algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*

Oleh :

Hikmah Jusadi

NIM : 09021181320059

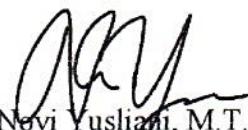
Indralaya, 05 Januari 2021

Pembimbing I



Dr. Abdiansah, S.Kom., M.Cs.  
NIP.198410012009121005

Pembimbing II



Novi Yusliani, M.T.  
NIP. 198211082012122001

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Rabu tanggal 23 Desember 2020 telah dilaksanakan Ujian Sidang Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Hikmah Jusadi

NIM : 09021181320059

Judul : Klasterisasi *Online Shop* Berdasarkan *Caption* Dengan Algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*

1. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs......  
NIP. 198410012009121005

2. Pembimbing II

Novi Yusliani, M.T......  
NIP. 198211082012122001

3. Pengaji I

Alvi Syahrini Utami, M.Kom......  
NIP. 197812222006042003

4. Pengaji II

Desty Rodiah, M.T......  
NIP.

Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Hikmah Jusadi  
NIM : 09021181320059  
Program Studi : Teknik Informatika  
Judul Skripsi : Klasterisasi *Online Shop* Berdasarkan *Caption* dengan Algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*

Hasil Pengecekan *Software iThenticate/Turnitine* : 14%

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan / plagiat. Apabila ditemukan unsur penjiplakan / plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, 05 Januari 2021

Hikmah Jusadi  
NIM. 09021181320059

## **MOTTO DAN PERSEMBAHAN**

- Kunci Dari Perjalanan Hidup Adalah Niat Yang Baik Dan Langkah Yang Kuat.
- Kebaikan Adalah Cara Terbaik Untuk Mengahadapi Setiap Keburukan.
- Positive Mind Produces Positive Things.
- You Are Who You Are, Be Proud And Keep Learning And Improve Everything To Be The Best Version Of Yourself.

### **Kupersembahkan Hasil Karyaku ini Kepada :**

- ❖ Allah SWT
- ❖ Kedua Orang Tuaku (Ibunda Ningsih Dan Ayahanda Alm. Herianto)
- ❖ Saudara-Saudariku Tercinta (Kak Herlina, Kak Hartini, Kak Sariani, Kak Sukma Prahayu, Kak Nirmala, Bang Herdiansyah Ainul Bahri, Dan Bang Hari Surya Pagi)
- ❖ IF Reguler 2013
- ❖ Alamamaterku
- ❖ Sahabatku Yang Selalu Ada Dalam Cerita Hidup Aku
- ❖ ‘Diri Sendiri’

# **ONLINE SHOP CLUSTERING BASED ON CAPTION USING JARO WINKLER DISTANCE AND K-NEAREST NEIGHBOR ALGORITHMS**

Hikmah Jusadi (09021181320059)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

## **ABSTRACT**

Captions on uploaded photos and videos on the online shop account in Instagram are not equipped with category filters, so the customers are often confused about determining the online shop category based on the caption. In this study the Jaro Winkler Distance and K-Nearest Neighbor algorithms were used to cluster 10 captions data from three different categories. Before the clustering process, pre-processing of text was carried out, namely case folding, tokenizing, and stopword removal which aimed to restore the standard form of text documents, break the document into words, and remove common words that has a high frequency of appearance. Jaro Winkler Distance is used to calculate the string similarity value between test data and training data and K-Nearest Neighbor is used for the clustering process based on predetermined K values, namely 3, 53, 103, 153, 203, 249, and 449. The K value with the highest accuracy result is K 249. Based on the analysis conducted from seven tests of 450 training data and 10 test data, the results obtained Precision 0.419, Recall 0.567, F-Measure 0.398, and 66% accuracy value.

*Keyword* : Clustering, Caption, Instagram, Jaro Winkler Distance, K-Nearest Neighbor

**KLASTERISASI ONLINE SHOP BERDASARKAN CAPTION DENGAN  
ALGORITMA JARO WINKLER DISTANCE DAN K-NEAREST  
NEIGHBOR**

Hikmah Jusadi (09021181320059)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

**ABSTRAK**

*Caption* pada unggahan foto dan video di akun *online shop instagram* belum dilengkapi dengan *filter* kategori, sehingga para konsumen atau pembeli masih sering kebingungan untuk menentukan kategori *online shop* berdasarkan *caption* tersebut. Dalam penelitian ini algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor* digunakan untuk melakukan klasterisasi 10 data *caption* dari tiga kategori yang berbeda. Sebelum proses klasterisasi, dilakukan proses prapengolahan teks, yaitu *case folding*, *tokenizing*, dan *stopword removal* yang bertujuan untuk mengembalikan bentuk standar dokumen teks, memecah dokumen menjadi kata-kata, dan pembuangan kata umum yang memiliki frekuensi kemunculan yang tinggi. *Jaro Winkler Distance* digunakan untuk menghitung nilai kesamaan *string* antara data uji dan data latih dan *K-Nearest Neighbor* digunakan untuk proses klasterisasi berdasarkan nilai K yang sudah ditentukan yaitu 3, 53, 103, 153, 203, 249, dan 449. Nilai K dengan hasil akurasi tertinggi adalah K 249. Berdasarkan analisis yang dilakukan dari tujuh kali pengujian dari 450 data latih dan 10 data uji mendapatkan hasil *Precision* 0,419, *Recall* 0,567, *F-Measure* 0,398, dan nilai *accuracy* 66%.

*Keyword* : Klasterisasi, *Caption*, *Instagram*, *Jaro Winkler Distance*, *K-Nearest Neighbor*

## KATA PENGANTAR

Segala puji dan syukur bagi Allah SWT atas berkah dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir yang berjudul “Klasterisasi *Online Shop* Berdasarkan *Caption* dengan Algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*” ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini Penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus Penulis ingin menyampaikan terima kasih kepada :

1. Allah SWT yang selalu mengabulkan doa hamba-Nya.
2. Diri Sendiri, Hikmah Jusadi yang sudah berhasil melewati beberapa cobaan selama penyelesaian tugas akhir ini.
3. Keluarga tercinta, Ibu yang selalu menyemangati di setiap langkahku dan selalu menyebut namaku didalam doanya, Alm. Ayah yang sudah tenang disana, dan 7 saudara-saudariku yang sudah menyemangati dan memberi dukungan moral dan materil selama menyelesaikan studi dan Tugas Akhir ini.
4. Ibu Nurida dan keluarga, sebagai ibu angkat dan keluarga disini, yang selalu mendukung dan mengingatkan untuk menyelesaikan studi dan Tugas Akhir ini.
5. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Dosen Pembimbing I dan Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing II yang telah membimbing, mengarahkan, dan memotivasi saya untuk menyelesaikan Tugas Akhir ini.
8. Ibu Alvi Syahrini Utami, M.Kom. dan Ibu Desty Rodiah, M.T. selaku penguji Tugas Akhir yang telah memberi nasihat dan saran yang membangun.
9. Bapak M. Fachrurrozi, M.T. selaku dosen pembimbing akademik yang selalu memberikan saran dan arahan selama masa perkuliahan.
10. Seluruh Dosen Program Studi Teknik Informatika yang telah memberikan bekal ilmu selama masa perkuliahan.
11. Sahabat seperjuangan Teknik Informatika Angkatan 2013.
12. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.

Penulis menyadari masih banyak terdapat kekurangan dalam penyusunan Tugas Akhir disebabkan oleh keterbatasan pengetahuan dan pengalaman. Oleh karena itu kritik dan saran membangun sangat diharapkan untuk menyempurnakan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Indralaya, 05 Januari 2021



Hikmah Jusadi

## DAFTAR ISI

	<b>Halaman</b>
<b>HALAMAN JUDUL.....</b>	.i
<b>LEMBAR PENGESAHAN TUGAS AKHIR.....</b>	ii
<b>TANDA LULUS UJIAN SIDANG TUGAS AKHIR .....</b>	iii
<b>HALAMAN PERNYATAAN.....</b>	iv
<b>MOTTO DAN PERSEMBAHAN.....</b>	v
<b>ABSTRACT.....</b>	vi
<b>ABSTRAK.....</b>	vii
<b>KATA PENGANTAR.....</b>	viii
<b>DAFTAR ISI .....</b>	x
<b>DAFTAR TABEL .....</b>	xiv
<b>DAFTAR GAMBAR.....</b>	xvii
<b>BAB I .....</b>	I-1
<b>PENDAHULUAN .....</b>	I-1
<b>1.1. Pendahuluan.....</b>	I-1
<b>1.2. Latar Belakang .....</b>	I-1
<b>1.3. Perumusan Masalah.....</b>	I-3
<b>1.4. Tujuan Penelitian .....</b>	I-4
<b>1.5. Manfaat Penelitian .....</b>	I-4
<b>1.6. Batasan Masalah .....</b>	I-4
<b>1.7. Sistematika Penulisan .....</b>	I-5
<b>1.8. Kesimpulan.....</b>	I-6
<b>BAB II.....</b>	II-1
<b>KAJIAN LITERATUR.....</b>	II-1
<b>2.1. Pendahuluan .....</b>	II-1
<b>2.2. Klasterisasi.....</b>	II-1
<b>2.3. Prapengolahan Teks.....</b>	II-2
<b>2.3.1. Case Folding.....</b>	II-2
<b>2.3.2. Tokenizing .....</b>	II-3

<b>2.3.3.</b>	<b><i>Stopword Removal</i></b> .....	II-4
<b>2.4.</b>	<b><i>Jaro-Winkler Distance</i></b> .....	II-5
<b>2.5.</b>	<b><i>K-Nearest Neighbor</i></b> .....	II-7
<b>2.6.</b>	<b><i>Instagram</i></b> .....	II-8
<b>2.7.</b>	<b>Pengujian <i>Confusion Matrix</i></b> .....	II-9
<b>2.7.1.</b>	<b><i>Precision</i></b> .....	II-10
<b>2.7.2.</b>	<b><i>Recall</i></b> .....	II-10
<b>2.7.3.</b>	<b><i>F-Measure</i></b> .....	II-10
<b>2.7.4.</b>	<b><i>Accuracy</i></b> .....	II-11
<b>2.8.</b>	<b>Penelitian Terkait</b> .....	II-11
<b>2.8.1.</b>	<b>Muhammad Rivki dan Adam Mukharil Bachtiar (2017)</b> .....	II-11
<b>2.8.2.</b>	<b>Anna Kurniawati, Sulistyo Puspitodjati, dan Sazali Rahman (2010)</b> .....	II-12
<b>2.8.3.</b>	<b>Yane Marita Febrianti, Indriati, dan Agus Wahyu Widodo (2018)</b> .....	II-12
<b>2.9.</b>	<b>Metode Pengembangan Perangkat Lunak</b> .....	II-13
<b>2.9.1.</b>	<b>Fase Insepsi</b> .....	II-13
<b>2.9.2.</b>	<b>Fase Elaborasi</b> .....	II-13
<b>2.9.3.</b>	<b>Fase Konstruksi</b> .....	II-14
<b>2.9.4.</b>	<b>Fase Transisi</b> .....	II-14
<b>2.10.</b>	<b>Kesimpulan</b> .....	II-15
<b>BAB III</b> .....		III-1
<b>METODOLOGI PENELITIAN</b> .....		III-1
<b>3.1.</b>	<b>Pendahuluan</b> .....	III-1
<b>3.2.</b>	<b>Unit Penelitian</b> .....	III-1
<b>3.3.</b>	<b>Pengumpulan Data</b> .....	III-1
<b>3.3.1.</b>	<b>Jenis dan Sumber Data</b> .....	III-1
<b>3.3.2.</b>	<b>Metode Pengumpulan Data</b> .....	III-2
<b>3.4.</b>	<b>Tahapan Penelitian</b> .....	III-2
<b>3.4.1.</b>	<b>Membuat Kerangka Kerja</b> .....	III-2
<b>3.4.2.</b>	<b>Menetapkan Kriteria Pengujian</b> .....	III-3
<b>3.4.3.</b>	<b>Format Data Pengujian</b> .....	III-7
<b>3.4.4.</b>	<b>Alat yang digunakan dalam Pelaksanaan Penelitian</b> .....	III-8

<b>3.4.5. Menganalisis Hasil Pengujian dan Membuat Kesimpulan.....</b>	<b>III-8</b>
<b>3.5. Managemen Proyek Penelitian .....</b>	<b>III-9</b>
<b>BAB IV.....</b>	<b>IV-1</b>
<b>PENGEMBANGAN PERANGKAT LUNAK.....</b>	
<b>4.1. Pendahuluan .....</b>	<b>IV-1</b>
<b>4.2. Fase Insepsi .....</b>	<b>IV-1</b>
<b>4.2.1. Pemodelan Bisnis.....</b>	<b>IV-2</b>
<b>4.2.2. Kebutuhan Sistem .....</b>	<b>IV-2</b>
<b>4.2.3. Analisis Dan Desain.....</b>	<b>IV-3</b>
<b>4.2.4. Desain Perangkat Lunak.....</b>	<b>IV-6</b>
<b>4.3. Fase Elaborasi.....</b>	<b>IV-15</b>
<b>4.3.1. Pemodelan Bisnis.....</b>	<b>IV-15</b>
<b>4.3.2. Kebutuhan Sistem .....</b>	<b>IV-17</b>
<b>4.3.3. Diagram Sequence .....</b>	<b>IV-17</b>
<b>4.4. Fase Kontruksi .....</b>	<b>IV-20</b>
<b>4.4.1. Kebutuhan Sistem .....</b>	<b>IV-20</b>
<b>4.4.2. Diagram Kelas .....</b>	<b>IV-21</b>
<b>4.4.3. Implementasi.....</b>	<b>IV-22</b>
<b>4.5. Fase Transisi .....</b>	<b>IV-24</b>
<b>4.5.1. Pemodelan Bisnis.....</b>	<b>IV-24</b>
<b>4.5.2. Kebutuhan Sistem .....</b>	<b>IV-25</b>
<b>4.5.3. Rencana Pengujian.....</b>	<b>IV-25</b>
<b>4.5.4. Implementasi.....</b>	<b>IV-26</b>
<b>4.6. Kesimpulan.....</b>	<b>IV-29</b>
<b>BAB V.....</b>	<b>V-1</b>
<b>HASIL DAN ANALISIS PENELITIAN .....</b>	
<b>5.1. Pendahuluan .....</b>	<b>V-1</b>
<b>5.2. Data Hasil Pengujian.....</b>	<b>V-1</b>
<b>5.2.1. Data Hasil Skenario I .....</b>	<b>V-2</b>
<b>5.2.2. Data Hasil Skenario II.....</b>	<b>V-3</b>
<b>5.2.3. Data Hasil Skenario III .....</b>	<b>V-5</b>

<b>5.2.4.</b>	<b>Data Hasil Skenario IV .....</b>	V-6
<b>5.2.5.</b>	<b>Data Hasil Skenario V.....</b>	V-8
<b>5.2.6.</b>	<b>Data Hasil Skenario VI .....</b>	V-9
<b>5.2.7.</b>	<b>Data Hasil Skenario VII.....</b>	V-11
<b>5.3.</b>	<b>Analisis Hasil Penelitian.....</b>	V-12
<b>5.4.</b>	<b>Kesimpulan .....</b>	V-14
<b>BAB VI.....</b>		VI-1
<b>KESIMPULAN DAN SARAN .....</b>		VI-1
<b>6.1.</b>	<b>Pendahuluan .....</b>	VI-1
<b>6.2.</b>	<b>Kesimpulan .....</b>	VI-1
<b>6.3.</b>	<b>Saran .....</b>	VI-2
<b>DAFTAR PUSTAKA .....</b>		xix
<b>LAMPIRAN I : KODE PROGRAM .....</b>		LI-1

## DAFTAR TABEL

Tabel II-1 <i>Case Folding</i> .....	II-3
Tabel II-2 <i>Tokenizing</i> .....	II-3
Tabel II-3 <i>Stopword Removal</i> .....	II-4
Tabel II-4 Data klasifikasi aktual dan prediksi pada <i>Confusion Matrix</i> .....	II-9
Tabel III-1. Rancangan Hasil Pengujian.....	III-7
Tabel III-2. Spesifikasi Kebutuhan Perangkat Keras dan Lunak.....	III-8
Tabel III-3. Uji Akurasi Terhadap Data Klasterisasi Dengan <i>Jaro Winkler Distance</i> dan <i>K-Nearest Neighbor</i> dengan Data Asli.....	III-9
Tabel III-4. Tabel Penjadwalan Penelitian dalam Bentuk WBS.....	III-10
Tabel IV-1. Kebutuhan Fungsional.....	IV-4
Tabel IV-2. Kebutuhan Non Fungsional.....	IV-4
Tabel IV-3. Definisi Aktor <i>Use Case</i> .....	IV-7
Tabel IV-4. Definisi <i>Use Case</i> .....	IV-8
Tabel IV-5. Skenario <i>Use Case Load Data</i> .....	IV-9
Tabel IV-6. Skenario <i>Use Case</i> Klasterisasi.....	IV-10
Tabel IV-7. Implementasi Kelas.....	IV-22
Tabel IV-8. Rencana Pengujian <i>Use Case Load Data</i> .....	IV-26
Tabel IV-9. Rencana Pengujian <i>Use Case</i> Klasterisasi .....	IV-26
Tabel IV-10. Pengujian <i>Use Case Load Data</i> .....	IV-27
Tabel IV-11. Pengujian <i>Use Case</i> Klasterisasi .....	IV-28
Tabel V-1. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario I.....	V-2

Tabel V-2. Confussion Matrix Pengujian Skenario I dengan K=3.....	V-3
Tabel V-3. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=3.....	V-3
Tabel V-4. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario II.....	V-4
Tabel V-5. Confussion Matrix Pengujian Skenario I dengan K=53.....	V-4
Tabel V-6. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=53.....	V-5
Tabel V-7. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario III.....	V-5
Tabel V-8. Confussion Matrix Pengujian Skenario I dengan K=103.....	V-6
Tabel V-9. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=103.....	V-6
Tabel V-10. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario IV.. ..	V-7
Tabel V-11. Confussion Matrix Pengujian Skenario I dengan K=153.....	V-7
Tabel V-12. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=153.....	V-8
Tabel V-13. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario V.....	V-8
Tabel V-14. Confussion Matrix Pengujian Skenario I dengan K=203.....	V-9
Tabel V-15. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=203.....	V-9
Tabel V-16. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario VI.....	V-10
Tabel V-17. Confussion Matrix Pengujian Skenario I dengan K=249.....	V-10
Tabel V-18. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=249.....	V-11
Tabel V-19. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario VII.....	V-11

Tabel V-20. Confussion Matrix Pengujian Skenario I dengan K=449.....	V-12
Tabel V-21. Performa Akurasi Klasterisasi Online Shop Berdasarkan Caption dengan K=449.....	V-12
Tabel V-22. Hasil Klasterisasi <i>Online Shop</i> Berdasarkan <i>Caption</i> Skenario I.....	V-13

## **DAFTAR GAMBAR**

Gambar III-1. Diagram Blok Tahapan Penelitian.....	III-3
Gambar III-2. Diagram Tahapan Perangkat Lunak.....	III-5
Gambar III-3. Diagram blok proses perhitungan kemiripan <i>string</i> dengan <i>Jaro-Winkler Distance</i> .....	III-6
Gambar III-4. Penjadwalan Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-15
Gambar III-5. Penjadwalan Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian.....	III-15
Gambar III-6. Penjadwalan untuk Menentukan Kriteria Pengujian.....	III-15
Gambar III-7. Penjadwalan Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase <i>Inception</i> .....	III-16
Gambar III-8. Penjadwalan Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase <i>Elaboration</i> .....	III-16
Gambar III-9. Penjadwalan Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase <i>Construction</i> .....	III-16
Gambar III-10. Penjadwalan Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian Fase <i>Transition</i> .....	III-17
Gambar III-11. Penjadwalan untuk Melakukan Pengujian Penelitian.....	III-17

Gambar III-12. Penjadwalan untuk Melakukan Hasil Pengujian dan Membuat Kesimpulan.....	III-17
Gambar IV-1. Contoh Data Latih.....	IV-5
Gambar IV-2. Contoh Data Uji.....	IV-6
Gambar IV-3. Diagram <i>Use Case</i> .....	IV-7
Gambar IV-4. Kelas Analisis <i>Load Data</i> .....	IV-12
Gambar IV-5. Kelas Analisis Klasterisasi.....	IV-12
Gambar IV-6. Diagram Aktifitas <i>Load Data</i> .....	IV-13
Gambar IV-7. Diagram Aktifitas Klasterisasi.....	IV-14
Gambar IV-8. Rancangan Antarmuka.....	IV-16
Gambar IV-9. Diagram <i>Sequence Load Data</i> .....	IV-18
Gambar IV-10. Diagram <i>Sequence Klasterisasi</i> .....	IV-19
Gambar IV-11. Diagram Kelas Perangkat Lunak.....	IV-21
Gambar IV-12. Tampilan Antarmuka Menu Utama.....	IV-24

## **BAB I**

### **PENDAHULUAN**

#### **1.1. Pendahuluan**

Pada bab ini membahas secara sistematis mengenai penelitian tugas akhir yang meliputi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah yang menjelaskan batasan perangkat lunak yang akan dibuat.

#### **1.2. Latar Belakang**

Seiring dengan pesatnya perkembangan dunia era *digital* saat ini, banyak pengguna internet yang melakukan jual beli melalui media sosial salah satunya *Instagram*. Menurut data *We Are Social* dan *Hootsuite*, hingga Januari 2020 pengguna *instagram* di Indonesia mencapai 63 juta orang (Febritania dan Riyanto, 2020). Melalui *instagram*, konsumen atau pembeli dapat mencari suatu barang atau produk dengan melihat unggahan iklan dari penjual berupa foto atau video dan melihat penjelasan mengenai barang tersebut dalam bentuk *caption* dengan maksimal karakter sebanyak 2200 karakter (Kalangit, 2016). *Caption* yang digunakan biasanya sangat variatif dengan kalimat bersifat pesuasi yang mampu membuat konsumen ingin membeli produk tersebut. Namun, pada bagian *caption* ini belum dilengkapi dengan *filter* kategori, sehingga para pembeli masih sering kebingungan untuk menentukan kategori online shop dari caption yang ada,

sehingga dibutuhkan suatu perangkat lunak yang dapat mengklasterisasi *online shop* berdasarkan *caption* tersebut.

Klasterisasi *online shop* dari *caption* dapat dilakukan dengan cara memproses setiap kata pada *caption* dan dikelompokkan berdasarkan kata kunci (*keyword*) yang mengacu pada kategori utama. Penelitian sebelumnya pernah dilakukan oleh Rivki dan Bachtiar (2017) mengenai klasifikasi *follower twitter* menggunakan *K-Nearest Neighbor (K-NN)* dengan kesimpulan bahwa *K-Nearest Neighbor* mampu mengklasifikasi *follower twitter* dengan akurasi 68% dari 4 pengujian. Selanjutnya, Febrianti, dkk (2018) melakukan penelitian menggunakan algoritma *K-Nearest Neighbor (K-NN)* dan *Jaro Winkler Distance* untuk kasus analisis sentimen pada ulasan salah satu aplikasi *online shop*. Kesimpulan dari penelitian ini menunjukan bahwa *K-Nearest Neighbor* dan *Jaro Winkler Distance* dapat diterapkan dan proses implementasi yang baik dengan hasil akurasi sebesar 74%.

*Jaro Winkler Distance* adalah algoritma yang digunakan untuk mencari nilai kesamaan antara dua *string* (Rochmawati dan Kusumaningrum, 2015). Algoritma ini dapat digunakan untuk mencocokkan *string* pada *caption instagram* dengan *string* yang sudah ditentukan yang akan menghasilkan nilai tertentu dari perbandingan kedua *string* tersebut. Kelebihan dari algoritma ini yaitu mampu mencari kesamaan *string* dengan *quadratic runtime complexity* yang sangat efektif (Kurniawati dkk, 2010).

*K-Nearest Neighbor* merupakan algoritma yang dapat mengklasifikasikan suatu objek berdasarkan jenis dan sampel yang digunakan (Lidya dkk, 2015).

Algoritma *K-NN* mampu mengklasterisasi *online shop* dengan kategori yang sudah ditentukan dalam waktu proses yang cepat dan memiliki tingkat akurasi yang tinggi.

Berdasarkan uraian diatas, penelitian ini akan mengembangkan sebuah perangkat lunak yang mampu mengklasterisasi jenis *online shop instagram* berdasarkan *caption* menggunakan algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*.

### 1.3. Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka penelitian tugas akhir yang dilakukan dalam penelitian ini adalah klasterisasi *online shop* berdasarkan *caption* dengan algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor*. Demi mendukung rumusan masalah tersebut, maka penelitian ini akan dibagi dalam *research questions*, yaitu:

1. Bagaimana mekanisme algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor* dalam klasterisasi *online shop* berdasarkan *caption*.
2. Bagaimana tingkat akurasi yang dihasilkan oleh algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor* dalam klasterisasi *online shop* berdasarkan *caption*.

#### **1.4. Tujuan Penelitian**

Adapun tujuan dari penelitian tugas akhir ini yaitu :

1. Mengetahui mekanisme algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor* dalam klasterisasi *online shop* berdasarkan *caption*.
2. Mengetahui tingkat akurasi yang dihasilkan oleh algoritma *Jaro Winkler Distance* dan *K-Nearest Neighbor* dalam klasterisasi *online shop* berdasarkan *caption*.

#### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian yaitu :

1. Hasil dari penelitian ini dapat mengklaster kategori *online shop* berdasarkan *caption* sehingga dapat digunakan untuk mengetahui kategori *online shop* dengan tepat
2. Hasil penelitian ini dapat digunakan dalam memilih algoritma yang tepat untuk klasterisasi pada penelitian-penelitian selanjutnya.

#### **1.6. Batasan Masalah**

Batasan masalah pada penelitian ini meliputi :

1. Data masukan adalah *caption instagram* dalam format \*.xlsx \*.txt.
2. Klasterisasi terdiri dari 3 kategori.
3. Data latih yang digunakan sebanyak 30 *caption instagram* dari 15 *online shop*.
4. Data uji yang digunakan sebanyak 10 *caption*.

## **1.7. Sistematika Penulisan**

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

### **BAB I. PENDAHULUAN**

Pada bab ini akan diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, Batasan masalah/ruang lingkup, metodologi penelitian dan sistematika penulisan.

### **BAB II. KAJIAN LITERATUR**

Pada bab ini akan membahas seluruh dasar-dasar teori yang digunakan mulai dari Prapengolahan, *Jaro Winkler Distance*, *K-Nearest Neighbor*, Pengujian Akurasi, Metode Pengembangan Perangkat Lunak dan semua yang digunakan pada tahapan analisis, perancangan, dan implementasi.

### **BAB III. METODOLOGI PENELITIAN**

Pada bab ini akan membahas mengenai tahap-tahap yang akan diterapkan pada penelitian. Setiap rencana dari tahapan penelitian dideskripsikan secara rinci berdasarkan kerangka kerja. Dilanjutkan dengan perancangan manajemen proyek dalam pelaksanaan penelitian.

### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan membahas mengenai alat yang digunakan dalam pelaksanaan penelitian. Pengembangan perangkat lunak berdasarkan panduan *Rational Unified Process(RUP)*. Terdapat empat fase pengembangan perangkat lunak pada *RUP* yaitu fase inisiasi, elaborasi, konstruksi, dan transisi.

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Pada bab ini akan membahas mengenai hasil dari pengujian dan tingkat akurasi perangkat lunak dan menganalisis hasil dari setiap pengujian untuk menentukan kesimpulan dari hasil pengujian yang dilakukan

## **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini akan membahas mengenai kesimpulan dari keseluruhan penelitian yang didapat dari Bab-bab sebelumnya dan juga saran untuk penelitian selanjutnya.

### **1.8. Kesimpulan**

Dari pendahuluan ini, telah dijelaskan secara umum mengenai penelitian yang akan dilakukan, meliputi latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan sehingga dilakukannya penelitian ini.

## **DAFTAR PUSTAKA**

- Denœux, T. (1995). A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(5), 804–813. <https://doi.org/10.1109/21.376493>
- Dwinavinta, D., Nugraha, C., Fahmi, M., Naimah, Z., & Setiani, N. (2014). Klasterisasi Judul Buku dengan Menggunakan Metode K-Means. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 21(1), 1907–5022.
- Dzikrulloh, N. N., & Setiawan, B. D. (2017). Penerapan Metode K – Nearest Neighbor ( KNN ) dan Metode Weighted Product ( WP ) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi ( Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri ). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(5), 378–385.
- Febrianti, Y. M., Indriati, I., & Widodo, A. W. (2018). Analisis Sentimen Pada Ulasan “Lazada ” Berbahasa Indonesia Menggunakan K-Nearest Neighbor ( K-NN ) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3689–3698.
- Febritania, G. R., & Sugeng Riyadi. (2020). Jurnal Indonesia Sosial Sains, 1(September), 132–141.

Insani, F. N. F. (2017). Rancang Bangun Perangkat Lunak Teenstagram Untuk Mengelompokkan Topik Caption Akun Intagram Siswa Sekolah Menengah Pertama Di Surabaya.

Kalangit, R.F., (2016). Alih Kode Dalam Instagram. *Jurnal Sastra Inggris*, 3345–3356.

Kornain, A., Yansen, F., & Tinaliah, T. (2014). Penerapan Algoritma Jaro-Winkler Distance Untuk Sistem Pendekripsi Plagiarisme Pada Dokumen Teks Berbahasa Indonesia. *Program Studi Teknik Informatika STMIK GI MDP*, 1–10. Retrieved from <http://eprints.mdp.ac.id/1068/>

Kurniawati, A., Puspitodjati, S., & Rahman, S. (2010). Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia. *Proceeding, Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen KOMMIT 2008, Depok, Indonesia*.

Kwon, O. W., & Lee, J. H. (2003). Text categorization based on k-nearest neighbor approach for Web site classification. *Information Processing and Management*, 39(1), 25–44. [https://doi.org/10.1016/S0306-4573\(02\)00022-5](https://doi.org/10.1016/S0306-4573(02)00022-5)

Leonardo, B., & Hansun, S. (2017). Text documents plagiarism detection using Rabin-Karp and Jaro-Winkler distance algorithms. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(2), 462–471.  
<https://doi.org/10.11591/ijeecs.v5.i2.pp462-471>

Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (Svm). *Seminar Nasional Teknologi Dan Komunikasi 2015*, 2015(Sentika), 1–8.  
<https://doi.org/10.1016/j.eswa.2013.08.047>

Min-Ling Zhang, & Zhi-Hua Zhou. (2005). A k-nearest neighbor based algorithm for multi-label classification, 718-721 Vol. 2.  
<https://doi.org/10.1109/grc.2005.1547385>

Mustakim, G. O. (2016). Algoritma K-Nearest Neighbor Classification. *Jurnal Sains, Teknologi Dan Industri*, 13(2), 195–202. Retrieved from <http://ejournal.uin-suska.ac.id/index.php/sitekin>

Parwita, W. G. S. (2019). Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 14(1), 27. <https://doi.org/10.30872/jim.v14i1.1272>

Pradnyana, G. A., & Djunaidy, A. (2013). Metode Weighted Maximum Capturing Untuk Klasterisasi Dokumen Berbasis Frequent Itemsets. *Jurnal Ilmu Komputer*, 6(2), 1–10.

Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31. <https://doi.org/10.21609/jsi.v13i1.500>

Rozaq, M. F., & Rozaq, M. F. (2019). Analisis : Jurnal Studi Keislaman Pengaruh Meme Terhadap Identitas Pemuda Muslim Nusantara, 19(1), 193–210.

Sebastian, D. (2017). Rancang Bangun Website Klasifikasi Untuk Pencarian Produk Pasar Online Menggunakan Algoritma K-Nearest Neighbor, 3, 417–432.

Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>

Suryaputra P., A., Samopa, F., & Hindayanto, B. C. (2014). Klasterisasi Dan Analisis Trafik Internet Menggunakan Fuzzy C Mean Dengan Ekstraksi Fitur Data. *Jurnal Informatika*, 12(1), 33–39. <https://doi.org/10.9744/informatika.12.1.33-39>

Tan, S. (2005). Neighbor-weighted K-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, 28(4), 667–671. <https://doi.org/10.1016/j.eswa.2004.12.023>

Tinaliah, T., & Elizabeth, T. (2018). Perbandingan Hasil Deteksi Plagiarisme Dokumen dengan Metode Jaro-Winkler Distance dan Metode Latent Semantic Analysis. *Jurnal Teknologi Dan Sistem Komputer*, 6(1), 7–12.  
<https://doi.org/10.14710/jtsiskom.6.1.2018.7-12>

Widiastuti, T. (2016). Rekayasa Gambar Diri Remaja dalam Mencapai Pengakuan Sosial di Instagram. *Jurnal Ilmu Komunikasi*, 14(2), 215–224.