

Perbandingan Metode *K-Nearest Neighbor* dan Metode *Naive Bayes*
untuk Klasifikasi Komentar Produk *Online Shop* Berdasarkan Ulasan
Pembeli

Diajukan Untuk Menyusun Tugas Akhir
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh :

Adi Nugroho

NIM : 09121002003

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018

LEMBAR PENGESAHAN USULAN TUGAS AKHIR

Perbandingan Metode *K-Nearest Neighbor* dan Metode *Naive Bayes*
untuk Klasifikasi Komentar Produk *Online Shop* Berdasarkan Ulasan
Pembeli

Oleh :

Adi Nugroho

NIM : 09121002003

Palembang, Januari 2019

Pembimbing I

Pembimbing II,



Rusdi Efendi, M.Kom
NIP 198201022011021201



Kanda Januar Miraswan, M.T
NIK 1671080901900006

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 13 Desember 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Adi Nugroho
NIM : 09121002003
Judul : Perbandingan Metode K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Komentar Produk Online Shop Berdasarkan Ulasan Pembeli

1. Ketua Penguji

Rusdi Efendi, M.Kom
NIP. 198201022011021201



2. Sekretaris

Kanda Januar Miraswan, M.T
NIK. 1671080901900006



3. Penguji I

Ir. M Ihsan Jambak, M.Sc., MM
NIP. 196804052013081201



4. Penguji II

Danny Matthew Saputra, M.Sc
NIP.198505102015041002



Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

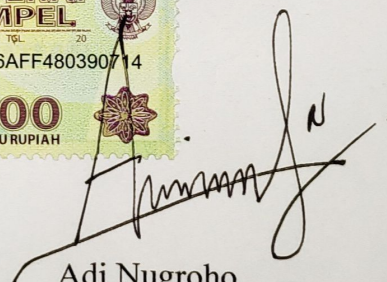
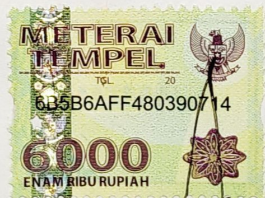
Yang bertanda tangan di bawah ini :

Nama : Adi Nugroho
NIM : 09121002003
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode K-Nearest Neighbor dan Metode Naive Bayes untuk Klasifikasi Komentar Produk Online Shop Berdasarkan Ulasan Pembeli
Hasil Pengecekan Software
iThenticate/Turnitin : 16 %

Menyatakan bahwa Laporan Projek 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, Januari 2019



Adi Nugroho
NIM. 09121002003

Motto :

- ❖ Maka bersabarlah kamu, karena sesungguhnya janji Allah itu benar, dan mohonlah ampunan untuk dosamu, serta bertasbihlah seraya memuji Tuhan-mu pada waktu petang dan pagi hari - Q.S. Al Mu'min : 55.
- ❖ Falling down is a part of life, getting back up is living - Self Reminder.
- ❖ An inch of time is an inch of gold, but you cant buy that inch of time with an inch of gold.
- ❖ Live the moment and don't think about future because we don't know what will happen the next day.
- ❖ Keep smile eventhough it's not easy but if you have something that you wanna reach, believe it, keep on your way, praying, and be strong - Ajeng Dwi Mastika.

Karya tulis ini didedikasikan kepada :

- Allah SWT
- Rasulullah Muhammad SAW
- Orang Tuaku
- Abang dan Kakak
- Sahabat-sahabatku Teknik Informatika 2012
- Pembimbing dan Penguji dari Universitas Sriwijaya
- Almamater, serta
- Teman-temanku

**PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN METODE NAIVE
BAYES UNTUK KLASIFIKASI KOMENTAR PRODUK ONLINE SHOP
BERDASARKAN ULASAN PEMBELI**

Adi Nugroho

09121002003

ABSTRAK

Klasifikasi adalah proses memprediksi kelas data atau objek yang label kelasnya belum diketahui ke label atau kelas yang sudah ada. Metode untuk klasifikasi diantaranya adalah *K-Nearest Neighbor* dan *Naive Bayes*, kedua metode tersebut dibandingkan untuk klasifikasi komentar yang dilakukan berdasarkan kata yang terkandung dalam komentar yang berdasarkan ulasan pembeli pada produk *online shop*. Hasil dari perbandingannya, metode *Naive Bayes* menghasilkan tingkat akurasi *recall* yang lebih baik dari *K-Nearest Neighbor* pada kelas negatif dengan melakukan uji *T Test* dengan nilai 0,78 sedangkan *recall K-Nearest Neighbor* pada kelas negatif bernilai 0,61.

Kata kunci: Metode *K-Nearest Neighbor*, Metode *Naive Bayes*, Klasifikasi Komentar.

**COMPARISON OF K-NEAREST NEIGHBOR METHOD AND NAÏVE BAYES
METHOD FOR CLASSIFICATION OF ONLINE SHOP PRODUCT COMMENT
BASED ON BUYER REVIEWS**

Adi Nugroho

09121002003

ABSTRACT

Classification is the process of predicting data class or object which class label is unknown to the existing label or class. The classification methods including K- Nearest Neighbor and Naive Bayes. Both methods are compared for the comments classification based on the words contained in the comments that based on buyer reviews on online shop product. The comparison result shows that Naïve Bayes obtained better recall accuracy rate than K-Nearest Neighbor in negative class using T-Test with value of 0.78, while recall value of K-Nearest Neighbor is 0.61.

Keywords: Comment Classification, K-Nearest Neighbor Method, Naive Bayes Method.

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT karena atas rahmat-Nya penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul “**Perbandingan Metode K-Nearest Neighbor dan Naive Bayes Untuk Klasifikasi Komentar Produk Online Shop Berdasarkan Ulasan Pembeli**” ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat S1 pada Jurusan Teknik Informatika Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada pihak-pihak telah memberikan dukungan, bimbingan dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, yaitu kepada:

1. Bapak dan Ibu tercinta HM Yunan dan Hj Zuraida. Abang dan Kakak yang selalu memberikan motivasi, do’a, saran, semangat;
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya;
3. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya;
4. Bapak Rusdi Efendi, M.Kom selaku dosen pembimbing I yang telah memberikan bimbingan, motivasi, saran serta bantuan dalam proses penyelesaian tugas akhir ini;
5. Bapak Kanda Januar Miraswan M.T selaku dosen pembimbing II yang telah memberikan bimbingan, motivasi, saran serta bantuan dalam proses penyelesaian tugas akhir ini;
6. Bapak Ir. M. Ihsan Jambak, M.Sc selaku dosen penguji I yang telah banyak memberikan saran dan kata-kata yang membangun;
7. Bapak Danny Matthew Saputra, S.T., M.Sc. selaku dosen penguji II yang telah banyak memberikan saran dan kata-kata yang membangun;
8. Bapak dan Ibu Dosen yang selama ini telah melimpahkan ilmunya kepada penulis selama proses belajar mengajar di Fakultas Ilmu Komputer Universitas Sriwijaya;

9. Partner ngoding dan bimbingan Agus Mistiawan, Leonard Sihaloho dan Riza Gamal Fuad yang telah banyak membantu selama pengerjaan tugas akhir sehingga bisa selesai dengan baik;
10. Staf akademik, kemahasiswaan, keuangan, dan pegawai yang selalu membantu dan mendukung penulis dalam hal kelancaran proses administrasi dan akademik selama masa perkuliahan;
11. Untuk semua pihak yang telah membantu dalam penyelesaian tugas akhir ini dan tidak dapat disebutkan satu-persatu;

Akhir kata, penulis menyadari bahwa tugas akhir ini jauh dari kata sempurna. Untuk itu penulis mengharapkan kritik dan saran yang membangun dari semua pihak untuk penyempurnaan tugas akhir ini dan semoga tugas akhir ini dapat bermanfaat bagi pihak yang membutuhkan.

Palembang, Januari 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
LEMBAR LULUS	iii
PERNYATAAN BEBAS PLAGIAT.....	iv
MOTTO	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi

BAB I PENDAHULUAN

1.1	Pendahuluan	I-1
1.2	Latar Belakang Masalah	I-1
1.3	Rumusan Masalah	I-2
1.4	Tujuan Penelitian	I-3
1.5	Manfaat Penelitian	I-3
1.6	Batasan Masalah	I-4
1.7	Sistematika Penulisan	I-4

BAB II TINJAUAN PUSTAKA

2.1	Pendahuluan.....	II-1
2.2	Landasan Teori.....	II-1
	2.2.1 Klasifikasi	II-1
	2.2.2 <i>Sentiment Analysis</i>	II-2
	2.2.3 <i>Preprocessing</i>	II-3

2.2.3.1	<i>Casefolding</i>	II-3
2.2.3.2	<i>Tokenizing</i>	II-4
2.2.3.3	<i>Stopword Removal</i>	II-4
2.2.3.4	<i>Stemming</i>	II-4
2.2.4	<i>Information Retrieval (Temu Kembali Informasi)</i>	II-7
2.2.4.1	<i>Term Frequency</i>	II-7
2.2.4.2	<i>Inverse Document Frequency (idf)</i>	II-8
2.2.4.3	Metode Pembobotan TF – IDF	II-8
2.2.5	<i>K-Nearest Neighbor</i>	II-10
2.2.6	<i>Naive Bayes</i>	II-15
2.2.7	<i>Cross Validation</i>	II-23
2.2.8	<i>Confusion Matrix</i>	II-24
2.2.9	<i>Rational Unified Process (RUP)</i>	II-25
2.3	Peneilitian Lain yang Relevan	II-28
2.4	Kesimpulan	II-29

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan	III-1
3.2	Metode Pengumpulan Data	III-1
3.2.1	Jenis Data	III-1
3.2.2	Sumber Data	III-1
3.2.3	Teknik Pengumpulan Data	III-2
3.3	Tahapan Penelitian	III-2
3.4	Diagram Blok Proses Umum Perangkat Lunak	III-3
3.5	Metode Pengembangan Perangkat Lunak	III-5
3.6	Teknik Pengujian	III-7
3.7	Penjadwalan Tugas Akhir	III-7

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1

4.2.1	Pemodelan Bisnis	IV-2
4.2.2	Kebutuhan	IV-2
4.2.3	Analisi	IV-2
4.2.4	Implementasi	IV-3
4.3	Fase Elaborasi	IV-4
4.3.1	Pemodelan Bisnis	IV-3
4.3.2	Kebutuhan	IV-12
4.3.3	Analisis dan Desain	IV-13
4.3.4	Implementasi	IV-18
4.4	Fase Konstruksi	IV-19
4.4.1	Pemodelan Bisnis	IV-19
4.4.2	Kebutuhan	IV-20
4.4.3	Analisis dan Desain	IV-21
4.4.4	Implementasi	IV-21
4.4.4.1	Implementasi Kelas	IV-21
4.4.4.2	Rancangan Antarmuka.....	IV-24
4.5	Fase Transisi	IV-25
4.5.1	Pemodelan Bisnis	IV-26
4.5.2	Kebutuhan	IV-26
4.5.3	Analisis dan Desain	IV-27
4.5.3.1	Rencana Pengujian <i>Use Case</i> memasukkan Data	IV-27
4.5.3.2	Rencana Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-28
4.5.3.3	Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i>	IV-28
4.5.4	Implementasi	IV-29
4.5.4.1	Pengujian <i>Use Case</i> Membuka File.....	IV-31
4.5.4.2	Pengujian Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-31

4.5.4.3 Pengujian Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i>	IV-31
--	-------

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Penelitian	V-1
5.2.1 Pembagian Data dengan <i>K-Fold Cross Validation</i>	V-1
5.2.2 Hasil Pengujian Metode <i>K-Nearest Neighbor</i>	V-2
5.2.3 Hasil Pengujian Metode <i>Naive Bayes</i>	V-13
5.2.4 <i>Confusion Matrix</i> Metode <i>K-Nearest Neighbor</i>	V-25
5.2.5 <i>Confusion Matrix</i> Metode <i>Naive Bayes</i>	V-26
5.2.6 nilai <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i> Metoder <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	V-26
5.2.7 Waktu Komputasi.....	V-27
5.3 Analisi Hasil Penelitian	V-28

BAB VI KESIMPULAN DAN SARAN

6.1 Kesimpulan	VI-1
6.2 Saran.....	VI-2

DAFTAR PUSTAKA	xvii
----------------------	------

DAFTAR GAMBAR

	Halaman
II-1. <i>Cross Validation</i>	II-24
II-2. Arsitektur RUP	II-26
III-1. Diagram Blok Proses Umum Perangkat Lunak <i>Sentiment Analysis</i> Menggunakan Perbandingan Metode <i>K-Neares Neighbor</i> dan Metode <i>Naive Bayes</i>	III-3
IV-1. <i>Use Case</i> Pengelompokan Secara Manual	IV-3
IV-2. Diagram <i>Use Case</i>	IV-6
IV-3. Kelas Analisis Mengelompokan Komentar <i>K-Nearest Neighbor</i>	IV-8
IV-4. Kelas Analisis Mengelompokan Komentar <i>Naive Bayes</i>	IV-8
IV-5. <i>Sequence Diagram</i> Membuka File	IV-9
IV-6. <i>Sequence Diagram K-Nearest Neighbor</i>	IV-10
IV-7. <i>Sequence Diagram Naive Bayes</i>	IV-11
IV-8. Kelas Diagram	IV-20
IV-9 Rancangan Antar Muka	IV-25
IV-10 <i>Mainform</i> Perangkat Lunak	IV-29
IV-11 Antarmuka Hasil Klasifikasi <i>K-Nearest Neighbor</i>	IV-30
IV-12 Antarmuka Hasil Klasifikasi <i>Naive Bayes</i>	IV-30
V-1 Perbandingan Nilai <i>Recall</i> Kelas Positif <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	V-29
V-2 Perbandingan Nilai <i>Recall</i> Kelas Negatif <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	V-29
V-3 Perbandingan Nilai <i>Recall</i> Kelas Netral <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	V-30
V-3 Perbandingan Waktu Komputasi <i>K-Nearest Neighbor</i> dan <i>Naive Bayes</i>	V-30

DAFTAR TABEL

	Halaman
II-1. Contoh Hasil <i>Preprocessing</i> Kalimat	II-6
II-2. Sampel Data Uji	II-12
II-3. Jumlah Term Pada Masing-Masing Dokumen	II-13
II-4. Jarak Dokumen Uji dengan Dokumen Latih	II-14
II-5. Urutan Kemiripan Dokumen	II-14
II-6. <i>Term</i> Dokumen <i>Testing</i> dan <i>Training</i>	II-17
II-7. Pembobotan <i>Term</i>	II-18
II-8. Bobot w setiap kelas	II-19
II-9. <i>Laplace Smoothing</i>	II-21
II-10. Proses <i>Matching</i>	II-21
II-11 <i>Confusion Matrix</i> untuk 2 kelas.....	II-24
III-1. Kegiatan Pengembangan Perangkat Lunak Berdasarkan RUP	III-6
III-2. Blok Penjadwalan Penelitian	III-8
IV-1 Definisi Aktor	IV-6
IV-2 Definisi <i>Use Case</i>	IV-7
IV-3 Kebutuhan Fungsional	IV-13
IV-4 Kebutuhan Non Fungsional	IV-13
IV-5 Skenario <i>Use Case</i> Membuka File	IV-15
IV-6 Skenario <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-17
IV-7 Skenario <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-17
IV-8 Daftar Implementasi Kelas	IV-22
IV-9 Rencana Pengujian <i>Use Case</i> Memasukkan Data	IV-27
IV-10 Rencana Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-28
IV-11 Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naïve Bayes</i>	IV-28
IV-12 Pengujian <i>Use Case</i> Membuka File.....	IV-32
IV-13 Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-32
IV-14 Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i>	IV-34
V-1 Pembagian Data Latih dan Data Uji	V-2

V-2	Klasifikasi Metode <i>K-Nearest Neighbor</i>	V-2
V-3	Hasil Klasifikasi Metode <i>K-Nearest Neighbor</i>	V-13
V-4	Klasifikasi Metode <i>Naive Bayes</i>	V-14
V-5	Hasil Klasifikasi Metode <i>Naive Bayes</i>	V-24
V-6	Nilai TP, FP, FN, dan TN <i>K-Nearest Neighbor</i>	V-25
V-7	Nilai TP, FP, FN, dan TN <i>Naive Bayes</i>	V-26
V-8	Nilai <i>Recall K-Nearest Neighbor</i>	V-27
V-9	Nilai <i>Recall Naive Bayes</i>	V-27
V-10	Lama Waktu Proses <i>K-Nearest Neighbor</i>	V-28
V-11	Lama Waktu Proses <i>Naive Bayes</i>	V-28

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2 Latar Belakang Masalah

Klasifikasi adalah proses penemuan fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006). Metode yang sering digunakan untuk klasifikasi diantaranya adalah metode *K-Nearest Neighbor* dan metode *Naive Bayes*. Kedua metode ini tergolong dalam *supervised learning*, yaitu pembelajaran terarah/terawasi. Proses pendekatannya sudah terdapat data yang dilatih, dan terdapat variabel yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. Namun pada penggunaannya *K-Nearest Neighbor* mengelompokkan objek berdasarkan data training yang paling dekat (mirip) dengan objek pada data baru atau data testing. Sedangkan metode *Naive Bayes* memanfaatkan teori probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya, tanpa menghitung

bobot dari data yang ada. *K-Nearest Neighbor* termasuk kedalam *lazy learning* yaitu tidak ada proses belajar data sedangkan *Naive Bayes* membutuhkan pembelajaran data untuk menentukan peluang.

Pada penelitian sebelumnya metode *K-Nearest Neighbor* dan metode *Naive Bayes* telah digunakan di beberapa penelitian seperti yang dilakukan oleh Westhyma Sibarani (2015) pada penelitiannya menggunakan metode *K-Nearest Neighbor* untuk klasifikasi citra batik besurek bengkulu dengan nilai akurasi 50% saat $k=3$. Dhanashree S. Medhekar, Mayur P. Bote, dan Shruti D. Deshmukh pada tahun 2013 menggunakan metode *Naive Bayes*. Penelitian ini menunjukkan *Naive Bayes* dapat digunakan sebagai klasifikasi untuk mendeteksi penyakit jantung. Hasil dari data training dan data testing menunjukkan akurasi klasifikasi hingga 89%.

Metode *K-Nearest Neighbor* dan *Naive Bayes* sudah digunakan untuk berbagai macam klasifikasi tetapi belum pernah digunakan untuk klasifikasi komentar pada produk *online shop*. Komentar produk sangat dibutuhkan sebagai rekomendasi kepada calon pembeli untuk mengetahui kualitas produk yang ingin dibelinya.

Berdasarkan uraian diatas penulis ingin melakukan penelitian perbandingan metode *K-Nearest Neighbor* dengan *Naive Bayes* untuk mengklasifikasikan komentar produk online shop berdasarkan ulasan pembeli.

1.3 Rumusan Masalah

Berdasarkan uraian-uraian pada latar belakang, rumusan masalah pada penelitian ini yaitu bagaimana perbandingan akurasi dan kecepatan metode *K-Nearest Neighbor* dan metode *Naive Bayes* mengklasifikasikan komentar dari ulasan pembeli.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Menerapkan metode *K-Nearest Neighbor* dan Metode *Naive Bayes* untuk klasifikasi komentar produk.
2. Mengetahui tingkat akurasi dan kecepatan dari metode *K-Nearest Neighbor* dan metode *Naive Bayes*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Sebagai bahan studi dan tambahan ilmu pengetahuan dalam hal klasifikasi teks atau dokumen bagi mahasiswa Fakultas Ilmu Komputer terutama Departemen Teknik Informatika yang ingin melakukan penelitian selanjutnya.
2. Sebagai masukan dan usulan kepada calon pembeli dalam melakukan klasifikasi komentar dari ulasan pembeli yang ada.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Ulasan pembeli yang diambil hanya dalam Bahasa Indonesia.
2. Data berupa dokumen yang berisi ulasan pembeli yang diambil dari *website online shop* Tokopedia.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian. Seperti definisi klasifikasi, *sentiment analysis*, *preprocessing*, *information retrieval*, *K-Nearest Neighbor*, *Naive Bayes* dan penelitian lain yang relevan dengan penelitian yang sedang dilakukan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini dibahas mengenai perancangan, lingkungan implementasi, dan hasil pengujian perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan pengembangan perangkat lunak klasifikasi komentar ini.

1.8 Kesimpulan

Berdasarkan uraian diatas, maka dalam penelitian ini akan dikembangkan sebuah perangkat lunak klasifikasi pada komentar produk berdasarkan ulasan pembeli Tokopedia menggunakan algoritma *K-Nearest Neighbor* dan *Naive Bayes*. Diharapkan metode yang digunakan dapat diimplementasikan ke dalam perangkat lunak.

DAFTAR PUSTAKA

- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika*. Bali, November 14, 2009.
- Alkhatib, Khalid. (2013). Stock Price Prediction Using K-Nearest Neighbor (KNN) Algorithm. *Internasional Journal of Business, Humanities and Technology*. Vol. 3, No. 3, March 2013.
- Brusilovsky, P., J.W. Ahn., E. Rasmussen. (2010). Teaching Information Retrieval with Web-Based Interactive Visualization. *Journal of Education for Library and Information Science* 51(3) : 187-200.
- Immanuel, Ricky. Ndaumanu. (2014). Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. Magister Teknik Informatika STMIK AMIKOM Yogyakarta. *Jurnal Teknik Informatika dan Sistem Informasi (JATISI)*, ISSN : 2407-4322, Vol. 1, No. 1, September 2014.
- Jiawei, H., Kamber, M., Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, *itd: Morgan Kaufmann*. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Kanagavalli, K., and S.T. Tharani. (2014). Analysing User Post for Web Forumusing K-Means Clustering. *International Journal of Scientific and Research Publications* 4(5).
- Kowalski, G. (2011). *Information Retrieval Architecture and Algorithms*. Springer, New York, USA.
- Kurniawan, H. (2012). Sistem Penentuan Kualitas Air pada Depot Ait Minum Menggunakan Metode K-Nearest Neighbor. Tugas Akhir Teknik Informatika. Pekanbaru: UIN Suska.
- Kusrini, Emha T. Luthfi, 2009, *Algoritma Data Mining*. Andi, Yogyakarta.
- Kruchten, P. (2000). *The Rational Unified Process an Introduction (2nd Edition)*. Addison Wesley. Boston, USA.
- Kwon, O. and Lee, J. (2003). Text categorization based on k-nearest neighbor approach for Web site classification. *Information Processing & Management*, 39(1), pp.25-44.
- Fitriah, Naila. 2012. Aplikasi Prediksi Harga Saham Apple, IBM, DELL, Dan HP Menggunakan Algoritma K-Nearest Neighbors. Teknik Informatika. Gunadarma University.

- Prilianti, Kestrilia Rega., and Wijaya, Hendra. (2014). Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. *Jurnal Cybermatika* Volume 2 Issue 1, 2014.
- Salton, G. (1971). *The SMART Retrieval System : Experiments in Automatic Document Processing*. Prentice Hall, USA.
- Saxena, Krati. (2014). Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm. *Internasional Journal of Computer Science Trends and Technology (IJCSST)* ISSN : 2347-8578 Volume 2 Issue 4, July-Agus 2014.
- Sukma, Alfian et-al . (2014). *K-Nearest Neighbor (Information Retrieval)*. Fakultas Sains dan Teknologi. Universitas Airlangga.
- Suyanto. (2008). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Informatika : Bandung.
- Sibarani, Westhyma. (2015). Implementasi Algoritma K-Nearest Neighbor Untuk Mengklasifikasi Motif Batik Besurek Bengkulu. Medan: Program Studi Teknologi Informasi Universitas Sumatera Utara.
- Tan, A. C., Naiman, D. Q., Xu, L., Winslow, R. L., & Geman, D. (2005). Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics*, 21(20), 3896–3904.
<https://doi.org/10.1093/bioinformatics/bti631>
- Widiarsana, O., Putra, N.W., Budiyasa, P.G.I., Bismantara, A.N.I., Mahajaya, S.N. (2011). *Data Mining: Metode Clasification K-Nearest Neighbor (KNN)*. Bali: Program Studi Teknologi Informasi Universitas Udayana.