

**KLASIFIKASI DOKUMEN MENGGUNAKAN WINNOWER
FINGERPRINT DENGAN METODE GAUSSIAN NAIVE BAYES**

Diajukan Untuk Menyelesaikan

Pendidikan Program Strata-I

Di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh :

Monica Shannaztamara Kanza

09021281320024

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

2019

LEMBAR PENGESAHAN TUGAS AKHIR

Klasifikasi Dokumen Menggunakan *Winnowing Fingerprint* Dengan Metode
Gaussian Naive Bayes

Oleh :

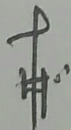
Monica Shannaztamara Kanza

NIM : 09021281320024

Pembimbing,

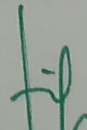
Mengetahui,

Ketua Jurusan Teknik Informatika



Yopy Sazaki, M.T.

NIP. 197406062015109101



Rifkie primartha, M.T.

NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 19 Desember 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Monica Shannaztamara Kanza

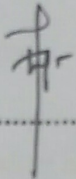
NIM : 09021281320024

Judul : Klasifikasi Dokumen Menggunakan *Winnowing Fingerprint* Dengan Metode *Gaussian Naive Bayes*.

1. Ketua Penguji

Yoppy Sazaki, M.T.

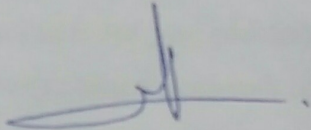
NIP. 197406062015109101


.....

2. Penguji I

Dr. Abdiansyah, S.Kom., M.Cs.

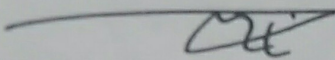
NIP. 198410012009121005


.....

3. Penguji II

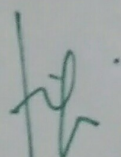
Osvari Arsalan, M.T.

NIP. 198806282018031001


.....

Mengetahui,

Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T.

NIP 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Monica S. Kanza
NIM : 09021281320024
Program Studi : Teknik Informatika
Judul Skripsi : Klasifikasi Dokumen Teks Menggunakan
Fingerprint Dengan Metode Gaussian Naive Bayes

Hasil Pengecekan Software *iThenticate/Turnitin* : 13%

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, 26 Desember 2019



Monica S. Kanza
NIM. 09021281320024

Motto dan Persembahan :

“Orang yang paling berbahagia di seluruh dunia bukanlah orang yang memiliki segala sesuatu yang terbaik, melainkan orang yang menjadikan yang terbaik dari segala sesuatu yang hadir dalam hidupnya.” Monica S Kanza

Kupersembahkan Karya Tulis ini kepada :

- Allah SWT
- Kedua orangtuaku
- Seluruh Keluargaku
- Kedua Pembimbing
- Seluruh Dosenku, serta
- Seluruh sahabatku

**KLASIFIKASI DOKUMEN MENGGUNAKAN *WINNOWERING*
FINGERPRINT DENGAN METODE *GAUSSIAN NAIVE BAYES***

OLEH :

MONICA S KANZA

09021281320024

ABSTRAKSI

Penelitian ini bertujuan untuk menganalisis kinerja algoritma Gaussian Naive Bayes dalam mengklasifikasikan *fingerprint* dokumen abstrak tugas akhir yang di ekstrak menggunakan algoritma *Winnowering Fingerprint*. Permasalahan yang akan dipecahkan pada penelitian ini yaitu seberapa besar akurasi Algoritma Gaussian Naive Bayes dalam mengklasifikasikan abstrak tugas akhir berdasarkan fingerprint-nya. Proses klasifikasi diawali dengan melakukan training dengan Algoritma Gaussian Naive Bayes terhadap fingerprint yang sudah di ekstrak dari abstrak tugas akhir dengan metode *winnowering fingerprint* untuk masing-masing kelas. Kemudian sejumlah data akan diuji untuk mengukur kinerja dari algoritma Gaussian Naive Bayes. Hasil dari penelitian menunjukkan bahwa algoritma Gaussian Naive Bayes kurang cocok untuk mengklasifikasikan dokumen berdasarkan fingerprint-nya. Hal ini terlihat dari nilai akurasi yang cukup rendah yaitu 33,2%.

**DOCUMENT CLASSIFICATION USING WINNOWING FINGERPRINT
WITH GAUSSIAN NAIVE BAYES METHOD**

BY :

MONICA S KANZA

09021281320024

ABSTRACT

The purpose of this research is to analyze the performance of Gaussian Naive Bayes Algorithm in classifying undergraduate thesis abstract based on its fingerprint which extracted using WInnowing Fingerprint Algorithm. The problem which aimed to be solved in this research is to find the accuration of Gaussian Naive Bayes algorithm in classifying undergraduate thesis abstract based o its fingerprint. The classification process begin with training a number of fingerprint data set, which class already known, using Gaussian Naive Bayes Algorithm for each class. Then, a number of new fingerprint data set will be tested to measure the performance of Gaussian Naive Bayes Algorithm. The result of this research shown that the percentage of the classification with the algorithm is 33,2%. Thus, we can conclude that gaussian Naive Bayes algorithm is not suitable to classify undergraduate thesis abstract data based on its fingerprint.

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT karena atas limpahan berkah dan karunia-Nya penulis dapat menyelesaikan tugas akhir ini. Tugas akhir ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat S1 di Jurusan Teknik Informatika Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tidak terhingga kepada pihak-pihak yang baik secara langsung maupun tidak langsung telah memberikan bantuan dan dukungan untuk menyelesaikan tugas akhir ini, antara lain :

1. Mama, Papa, dan adikku Faris atas semua bantuan, kasih sayang, dukungan dan doa yang tiada hentinya untuk penulis;
2. Saskia, Bella, Aldi, Tamma dan seluruh anggota keluarga, yang tidak dapat dituliskan satu persatu, atas semua bantuan, kasih sayang, serta dukungan mereka;
3. Putra Jaya S.Kom yang telah memberikan semangat dan dukungan selama masa sulit penulis dalam penulisan tugas akhir;
4. Bapak Jaidan Jauhari, S.Pd., M.T. selaku dekan Fakultas Ilmu Komputer Universitas Sriwijaya;
5. Bapak Rifkie Primartha, M.T. selaku kepala Jurusan Teknik Informatika Universitas Sriwijaya;
6. Ibu Hardini Novianti, M.T. selaku Sekretaris Jurusan Teknik Informatika Universitas Sriwijaya;

7. Bapak Yoppy Sazaki, M.T. selaku pembimbing 1 yang telah banyak membantu, mengajarkan, dan senantiasa bersabar dalam membimbing penulis;
8. Bapak Hadipurnawan Satria, Ph.D selaku pembimbing 2 yang telah banyak membantu dan membimbing penulis;
9. Ibu Novi Yusliani, M.T. selaku pembimbing akademik dan penguji;
10. Bapak Osvari Arsalan, M.T., Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku penguji;
11. Bapak dan Ibu dosen Fakultas Ilmu Komputer Universitas Sriwijaya yang selama ini telah melimpahkan ilmunya kepada penulis baik di dalam maupun di luar kegiatan belajar mengajar;
12. Kak Ricy Firnando selaku administrasi Fakultas Ilmu Komputer Jurusan Teknik Informatika Reguler, atas bantuan dan kesabarannya mengurus seluruh kegiatan perkuliahan;
13. Staf dan Karyawan Fakultas Ilmu Komputer Jurusan Teknik Informatika, atas bantuannya dalam memperlancar kegiatan akademik dan persyaratan Tugas Akhir;
14. Ade Rahma Youanda S.Kom, Septa Nopita Sari S.Kom, Dian Lukita Sari S.Kom, Faris Harun Ahmad S.Kom, Dian Lestari S.Kom, Dewi Putri Siagian S.Kom, Annisa Septi Anugrah S.Kom, serta seluruh teman-teman yang lain atas persahabatan dan dukungannya;
15. Semua pihak yang telah membantu dalam penyelesaian tugas akhir ini, yang tidak dapat dituliskan satu persatu.

Akhir Kata, penulis sepenuhnya menyadari bahwa tugas akhir ini sangat 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, 26 Desember 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN BEBAS PLAGIARISME	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRAKSI	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN.....	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	I-1
1.2 Perumusan Masalah.....	I-3
1.3 Tujuan Penelitian.....	I-3
1.4 Manfaat Penelitian.....	I-3
1.5 Batasan Masalah.....	I-4
1.6 Metodologi Penelitian	I-4
1.6.1 Unit Penelitian.....	I-4
1.6.2 Metode Pengumpulan Data	I-4
1.6.2.1 Jenis Data	I-4

1.6.2.2 Sumber Data.....	I-4
1.6.2.3 Teknik Pengumpulan Data.....	I-5
1.6.3 Langkah-langkah Penelitian.....	I-5
1.6.4 Metode Pengembangan Perangkat Lunak	I-6
1.7 Sistematika Penulisan.....	I-7
 BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait	II-1
2.2 Landasan Teori	II-2
2.2.1 Pemrosesan Bahasa Alami	II-2
2.2.2 Klasifikasi.....	II-3
2.2.3 Gaussian Naïve Bayes	II-3
2.2.4 Algoritma Wininging Fingerprint.....	II-8
2.2.4.1 Algoritma Fingerprint.....	II-8
2.2.4.2 Algoritma Wininging Fingerprint	II-10
2.2.5 Rational Unified Process	II-12
 BAB III ANALISIS DAN PERANCANGAN	
3.1 Analisis Masalah	III-1
3.2 Analisis Data	III-2
3.3 Analisis Proses Perangkat Lunak	III-2
3.3.1 Analisis Preprocessing	III-5
3.3.1.1 Analisis Tokenizing dan Case Folding	III-5
3.3.1.2 Analisis Stopword Removal	III-6
3.3.1.3 Analisis Stemming.....	III-6

3.3.2 Analisis Document Fingerprinting	III-7
3.3.2.1 Pembentukan K-Gram	III-8
3.3.2.2 Perhitungan Nilai Hash	III-8
3.3.2.3 Pembentukan Window	III-8
3.3.2.4 Seleksi Fingerprint	III-9
3.3.3 Analisis Klasifikasi	III-9
3.3.3.1 Perhitungan Nilai Mean	III-10
3.3.3.2 Perhitungan Nilai Standar Deviasi	III-10
3.3.3.3 Perhitungan Probabilitas Kelas	III-11
3.4 Analisis Perangkat Lunak	III-11
3.4.1 Analisis Kebutuhan Perangkat Lunak	III-11
3.4.1.1 Deskripsi Umum Perangkat Lunak	III-11
3.4.1.2 Kebutuhan Perangkat Lunak	III-12
3.4.1.3 Model Use Case	III-13
3.4.1.3.1 Diagram Use Case	III-14
3.4.1.3.2 Tabel Definisi Aktor dan Use Case	III-15
3.4.1.3.3 Skenario Use Case	III-17
3.4.1.4 Activity Diagram	III-22
3.4.1.5 Sequence Diagram	III-26
3.4.1.6 Kelas Diagram	III-31
3.4.1.7 Antarmuka Pengguna	III-32

BAB IV IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Perangkat Lunak	IV-1
--	------

4.1.1 Lingkungan Implementasi.....	IV-1
4.1.2 Implementasi Kelas	IV-2
4.1.3 Implementasi Antarmuka	IV-3
4.2 Pengujian Perangkat Lunak.....	IV-4
4.2.1 Metode Pengujian.....	IV-4
4.2.2 Lingkungan Pengujian.....	IV-5
4.2.3 Rencana Pengujian	IV-5
4.2.4 Kasus Uji.....	IV-8
4.2.5 Hasil Pengujian Perangkat Lunak	IV-17
4.3 Analisis Hasil Pengujian	IV-22
4.3.1 Hasil Pengujian Klasifikasi	IV-22
4.3.2 Hasil Pengujian Akurasi.....	IV-25
 BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan.....	V-1
5.2 Saran.....	V-2
DAFTAR PUSTAKA	VI

DAFTAR TABEL

Tabel 3.1 Contoh Hasil Proses <i>Case Folding</i> dan <i>Tokenizing</i>	III-6
Tabel 3.2 Contoh <i>Stopword Removal</i>	III-6
Tabel 3.3 <i>Fingerprint</i> Contoh Data.....	III-57
Tabel 3.4 Tabel Definisi Aktor	III-66
Tabel 3.5 Tabel Definisi <i>Use Case</i>	III-66
Tabel 3.6 Skenario <i>Use Case Input Data</i> dan Pilih Proses	III-68
Tabel 3.7 Skenario <i>Use Case Training</i>	III-69
Tabel 3.8 Skenario <i>Use Case Testing</i>	III-70
Tabel 3.9 Skenario <i>Use Case Preprocessing</i>	III-71
Tabel 3.10 Skenario <i>Use Case Ekstraksi Fingerprint</i>	III-72
Tabel 4.1 Daftar Implementasi Kelas.....	IV-2
Tabel 4.2 Rencana Pengujian <i>Use Case Input Data</i>	IV-6
Tabel 4.3 Rencana Pengujian <i>Use Case Training</i>	IV-7
Tabel 4.4 Rencana Pengujian <i>Use Case Testing</i>	IV-7
Tabel 4.5 Rencana Pengujian <i>Use Case Preprocessing</i>	IV-8
Tabel 4.6 Rencana Pengujian <i>Use Case Ekstraksi Fingerprint</i>	IV-8
Tabel 4.7 Pengujian <i>Use Case Input Data</i>	IV-9
Tabel 4.8 Pengujian <i>Use Case Training</i>	IV-11
Tabel 4.9 Pengujian <i>Use Case Testing</i>	IV-14
Tabel 4.10 Pengujian <i>Use Case Preprocessing</i>	IV-16
Tabel 4.11 Pengujian <i>Use Case Ekstraksi Fingerprint</i>	IV-17

Tabel 4.12 Hasil Pengujian Klasifikasi Dokumen IV-23

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Rational Unified Process.....	II-15
Gambar 3.1 Diagram Alur Proses Klasifikasi dengan Metode Naive Bayes Berdasarkan Algoritma Fingerprint.....	III-3
Gambar 3.2 Preprocessing	III-5
Gambar 3.3 Flowchart <i>Document Fingerprinting</i> dengan <i>Winnowing Fingerprint</i>	III-7
Gambar 3.4 Flowchart Klasifikasi dengan Algoritma <i>Naive Bayes</i>	III-9
Gambar 3.5 Diagram <i>Use Case</i> Sistem Pengklasifikasi Dokumen Teks Berbahasa Indonesia Menggunakan Naive Bayes Berdasarkan Algoritma <i>Winnowing Fingerprint</i>	III-65
Gambar 3.6 <i>Activity Diagram</i> Input Data	III-73
Gambar 3.7 <i>Activity Diagram</i> Proses <i>Training</i>	III-74
Gambar 3.8 <i>Activity Diagram</i> Proses <i>Testing</i>	III-74
Gambar 3.9 <i>Activity Diagram</i> Ekstraksi Fingerprint	III-75
Gambar 3.10 <i>Activity Diagram</i> <i>Preprocessing</i>	III-76
Gambar 3.11 Kelas Analisis <i>Input Data</i>	III-77
Gambar 3.12 Kelas Analisis <i>Training</i>	III-77
Gambar 3.13 Kelas Analisis <i>Testing</i>	III-77
Gambar 3.14 Kelas Analisis <i>Preprocessing</i>	III-78
Gambar 3.15 Kelas Analisis Ekstraksi <i>Fingerprint</i>	III-78
Gambar 3.16 <i>Sequence Diagram</i>	III-79

Gambar 3.17 Kelas Diagram.....	III-81
Gambar 3.18 Perancangan Antarmuka HomeInterface	III-82
Gambar 3.19 Perancangan Antarmuka TrainOutputInterface	III-83
Gambar 3.20 Perancangan Antarmuka TestOutputInterface	III-84
Gambar 4.1 Antarmuka <i>HomeInterface</i>	IV-3
Gambar 4.2 Antarmuka TestOutputInterface.....	IV-3
Gambar 4.3 Antarmuka TrainOutputInterface.....	IV-4
Gambar 4.4 Tampilan Utama Antarmuka Perangkat Lunak.....	IV-18
Gambar 4.5 Tampilan Jendela Pencarian File.....	IV-19
Gambar 4.6 Tampilan Utama Perangkat Lunak Setelah Pemilihan Folder ...	IV-20
Gambar 4.7 Tampilan Perangkat Lunak Ketika Radio Button Training Dipilih.....	IV-20
Gambar 4.8 Tampilan Hasil <i>Training</i>	IV-21
Gambar 4.9 Tampilan Hasil <i>Testing</i>	IV-22

DAFTAR LAMPIRAN

Lampiran 1 Tabel WBS Penjadwalan Penelitian	L1-1
Lampiran 2 <i>Coding</i> Program.....	L2-1

BAB I

PENDAHULUAN

1.1 Latar Belakang

Fakultas Ilmu Komputer merupakan salah satu fakultas di Universitas Sriwijaya. Untuk menyelesaikan studinya, mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya diwajibkan mengerjakan dan menyelesaikan Tugas Akhir berupa Skripsi. Seiring dengan berjalannya waktu, Tugas Akhir yang telah dikerjakan mahasiswa semakin banyak. Dokumen Tugas akhir yang semakin banyak belum terorganisir dengan baik secara digital, yang mengakibatkan sulitnya mahasiswa untuk mencari Tugas Akhir dari alumni sebagai referensi untuk menyelesaikan Tugas Akhir. Untuk mengorganisir dokumen Tugas Akhir mahasiswa secara digital diperlukan proses klasifikasi.

Salah satu penelitian terkait klasifikasi dilakukan oleh Yusra, Dhita Olivita, dan Yelfi Vitriani (2016). Penelitian dilakukan untuk membandingkan metode *Naive Bayes* dan *K-Nearest Neighbour* untuk mengklasifikasikan Tugas Akhir mahasiswa. Hasil dari penelitian ini menunjukkan bahwa *Naive Bayes* memiliki akurasi lebih baik dibandingkan dengan *K-Nearest Neighbour*. Hasil dari penelitian ini adalah *Naive Bayes* memiliki akurasi yang lebih baik dalam mengklasifikasikan data Tugas Akhir mahasiswa daripada *K-Nearest Neighbour*, dimana akurasi *Naive Bayes* sebesar 87% sedangkan *K-Nearest Neighbour* 86% untuk nilai $k=5$.

Untuk meningkatkan kecepatan proses klasifikasi, dapat dilakukan proses *Document Fingerprinting* terhadap dokumen teks, dimana dokumen teks tersebut akan diproses sehingga didapatkan suatu hasil berupa *fingerprint* atau sidik jari dari

suatu dokumen yang berfungsi sebagai ciri dan identitas dari dokumen tersebut. *Fingerprint* dari suatu dokumen didapatkan dengan cara menghitung nilai *hash* dari dokumen (Sanjaya & Absar, 2015).

Penelitian terkait *Document Fingerprinting* pernah dilakukan oleh Suwanto Sanjaya dan Ersad Alfarisy Absar (2015) untuk menguji akurasi pengklasifikasian dokumen teks berbasis *Winnowing Fingerprint* menggunakan algoritma *K-Nearest Neighbour* (KNN). *Fingerprint* yang di dapat dari hasil perhitungan algoritma *Winnowing Fingerprint* kemudian diproses untuk mendapatkan nilai tetangga terdekat. Dari penelitian yang dilakukan Suwanto Sanjaya dan Ersad Alfarisy Absar, dapat disimpulkan bahwa klasifikasi dengan menggunakan algoritma KNN untuk data yang *fingerprint*-nya didapat dengan algoritma *Winnowing Fingerprint* memiliki akurasi sebesar 80%.

Algoritma *Gaussian Naive Bayes* sudah sering digunakan dan terbukti baik untuk mengklasifikasikan dokumen. Namun algoritma ini belum pernah digunakan untuk dokumen yang telah di *hash* menggunakan algoritma *Winnowing Fingerprint* untuk dokumen Tugas Akhir berbahasa Indonesia. Maka dari itu, akan diteliti mengenai klasifikasi Tugas Akhir berbahasa Indonesia berbasis *Winnowing Fingerprint* menggunakan algoritma *Gaussian Naive Bayes*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan sebelumnya, maka dapat disimpulkan masalah yang akan diangkat dan diteliti pada penelitian ini adalah seberapa besar akurasi algoritma *Gaussian Naive Bayes* untuk melakukan proses klasifikasi terhadap dokumen teks berdasarkan algoritma *Winnowing Fingerprint*.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu :

1. Untuk mengetahui tingkat akurasi algoritma *Gaussian Naive Bayes* dalam melakukan proses klasifikasi terhadap dokumen berdasarkan *fingerprint* yang didapat dari hasil perhitungan algoritma *Winnowing Fingerprint*.

1.4 Manfaat Penelitian

Adapun manfaat yang akan didapat dari penelitian ini yaitu :

1. Dapat memahami cara penerapan algoritma *Gaussian Naive Bayes* untuk melakukan proses klasifikasi dokumen berdasarkan *fingerprint* yang didapat dari hasil perhitungan algoritma *Winnowing Fingerprint*
2. Dapat mengetahui seberapa besar akurasi algoritma *Gaussian Naive Bayes* dalam melakukan proses klasifikasi dokumen teks berdasarkan *fingerprint* dokumen yang didapatkan dengan algoritma *Winnowing Fingerprint*.
3. Hasil penelitian ini dapat dijadikan salah sebagai rujukan untuk penelitian selanjutnya.

1.5 Batasan Masalah

Batasan Masalah pada penelitian ini adalah sebagai berikut :

1. Proses klasifikasi dilakukan pada dokumen berbahasa Indonesia berdasarkan hasil ekstraksi *Fingerprint* dokumen.
2. Data yang digunakan dalam penelitian ini adalah dokumen abstrak Tugas Akhir mahasiswa dengan format txt.

3. Data yang diklasifikasikan terbagi menjadi lima kelas, yaitu Pengolahan Citra, Pemrosesan Bahasa Alami, Sistem Pakar, Sistem Informasi, dan Sistem Pendukung Keputusan.

1.6 Metodologi Penelitian

1.6.1 Unit Penelitian

Unit Penelitian yang digunakan pada penelitian ini adalah Fakultas Ilmu Komputer Universitas Sriwijaya.

1.6.2 Metode Pengumpulan Data

1.6.2.1 Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu dokumen teks berupa abstrak dari tugas akhir mahasiswa dalam format txt.

1.6.2.2 Sumber Data

Data penelitian yang digunakan pada penelitian ini adalah abstrak Tugas Akhir yang diperoleh dari Universitas Sriwijaya.

1.6.2.3 Teknik Pengumpulan Data

Data yang di gunakan pada penelitian ini di kumpulkan dengan cara menghubungi dan meminta izin untuk menggunakan data salah satu alumni Fakultas Ilmu Komputer Universitas Sriwijaya yang pernah melakukan penelitian mengenai klasifikasi.

1.6.3 Langkah-langkah Penelitian

Dalam melakukan penelitian ini, terdapat tahapan-tahapan yang harus dilalui, yaitu :

1. Mempelajari konsep algoritma *Gaussian Naive Bayes* dan *Winnowing Fingerprint*.
2. Melakukan studi literatur pada jurnal, buku, maupun thesis yang memuat penelitian sebelumnya mengenai *Gaussian Naive Bayes* dan *Winnowing Fingerprint*.
3. Mengumpulkan dokumen teks berupa abstrak dari Tugas Akhir yang akan dijadikan sebagai data uji maupun data latih.
4. Melakukan pengembangan perangkat lunak pengklasifikasi dokumen Tugas Akhir berdasarkan abstraknya dengan metode RUP.
5. Melakukan tahapan-tahapan penelitian yang telah dirancang untuk menemukan kelas dari Tugas Akhir secara tepat.
6. Melakukan analisis terhadap penelitian yang telah dilakukan.

1.6.4 Metode Pengembangan Perangkat Lunak

Metode yang digunakan dalam penelitian tugas akhir mengenai klasifikasi dokumen Tugas Akhir berdasarkan abstraknya ini adalah metode Rational Unified Poces (RUP). Adapun tahapan-tahapan RUP dalam penelitian ini yaitu sebagai berikut :

1. Fase Insepsi (*Inception*)

Pada fase ini, dilakukan beberapa hal, yaitu :

- a. Menentukan tahapan proses yang akan dilakukan oleh perangkat lunak mulai dari pembentukan *fingerprint* sampai dengan terdefinisinya kelas dari Tugas Akhir.
- b. Menentukan kebutuhan perangkat lunak seperti bagaimana input dokumen abstrak Tugas Akhir ataupun output kelas yang sudah ditentukan.

2. Fase Elaborasi (*Elaboration*)

Pada fase ini, dilakukan beberapa hal, yaitu :

- a. Melakukan analisis dan perancangan perangkat lunak pengklasifikasi dokumen Tugas Akhir berdasarkan abstraknya.
- b. Mengidentifikasi arsitektur perangkat lunak berdasarkan *use-case* yang telah terbentuk pada fase insepsi.
- c. Mengidentifikasi model kelas analisis, diagram kelas, *sequence diagram* dan diagram aktifitas dari perangkat lunak pengklasifikasi dokumen Tugas Akhir berdasarkan abstraknya.

3. Fase Konstruksi (*Construction*)

Pada Fase ini dilakukan beberapa hal, yaitu :

- a. Memastikan kesesuaian dan kelengkapan Antara diagram *use-case*, model kelas analisis, kelas diagram, dan *sequence diagram*.
- b. Membuat kode program yang sesuai dengan fungsi-fungsi yang telah digambarkan pada fase sebelumnya.
- c. Melakukan pengujian perangkat lunak dan perbaikan berdasarkan hasil analisis pengujian.

4. Fase Transisi (*Transition*)

Pada fase ini akan dilakukan pengujian terhadap perangkat lunak dengan metodologi pengujian perangkat lunak, yaitu metode *black box testing*.

1.7 Sistematika Penulisan

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

1. BAB I PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan, dan manfaat penelitian, metode penelitian, metode pengembangan perangkat lunak dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Pada bab ini diuraikan mengenai tinjauan pustaka berupa dasar-dasar teori yang akan digunakan pada analisis, perancangan, dan implementasi tugas akhir.

3. BAB III ANALISIS DAN PERANCANGAN

Pada bab ini diuraikan mengenai analisis dan perancangan terhadap penggunaan metode *Winnowing Fingerprint* dan *Gaussian Naive Bayes* yang dikembangkan menjadi perangkat lunak pengklasifikasi dokumen teks.

4. BAB IV IMPLEMENTASI DAN PENGUJIAN

Pada bab ini diuraikan mengenai lingkungan implementasi, implementasi program, hasil eksekusi, dan hasil uji terhadap perangkat lunak pengklasifikasi yang telah dibangun.

5. BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan terhadap hasil penelitian dan saran untuk pengembangan selanjutnya.

Daftar Pustaka

- Elbegbayan, N. (2005). Winnowing, a Document Fingerprinting Algorithm. *TDDC03 Projects, Spring*.
- <http://hikaruyuuki.lecture.ub.ac.id/kamus-kata-dasar-dan-stopword-list-bahasa-indonesia/> (terakhir di akses : 20-6-2019)
- Jain, A., & Mishra, R. D. (2016). Text Categorization : By Combining Naive Bayes and Maximum Entropy Classifiers. *International Journal of Advances in Electronics and Computer Science*.
- Klon, A. E., Lowrie, J. F., & Diller, D. J. (2006). Improved naive Bayesian modeling of numerical data for absorption, distribution, metabolism and excretion (ADME) property prediction. *Journal of chemical information and modeling*, 46(5), 1945-1956.
- Kruchten, P. (2004). *The rational unified process: an introduction*: Addison-Wesley Professional.
- Neelamegam, S., & Ramaraj, E. (2013). Classification algorithm in data mining: An overview. *International Journal of P2P Network Trends and Technology (IJPTT)*, 4(8), 369-374.
- Sanjaya, S., & Absar, E. A. (2015). Pengelompokan Dokumen Menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour. *Jurnal CoreIT*, 1(2), 50-56.
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *JURNAL TEKNIK INFORMATIKA DAN SISTEM INFORMASI (JATISI)*, VOLUME 3 151.

- Suciadi, J. (2004). Studi analisis metode-metode parsing dan interpretasi semantik pada natural language processing. *Jurnal Informatika*, 2(1), pp. 13-22.
- Surahman, A. M. (2013). Perancangan Sistem Penentuan similarity kode program bahasa C dan Pascal Dengan Menggunakan Algoritma Rabin-Karp.
- Wibowo, A. T., Sudarmadi, K. W., & Barmawi, A. M. (2013). *Comparison between fingerprint and winnowing algorithm to detect plagiarism fraud on Bahasa Indonesia documents*. Paper presented at the Information and Communication Technology (ICoICT), 2013 International Conference of.