

**ANALISIS PERBANDINGAN KLASIFIKASI DOKUMEN SKRIPSI
MENGUNAKAN SUPPORT VECTOR MACHINE, NAÏVE BAYES
CLASIFIER, DAN K-NEAREST NEIGHBOR
SKRIPSI**

Diajukan untuk memenuhi salah satu syarat S-1

Program Studi Teknik Informatika



Disusun oleh:

SD Abdurrahman

18106050029

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2022



PENGESAHAN TUGAS AKHIR

Nomor : B-1869/Un.02//PP.00.9/08/2022

Tugas Akhir dengan judul : ANALISIS PERBANDINGAN KLASIFIKASI DOKUMEN SKRIPSI MENGGUNAKAN SUPPORT VECTOR MACHINE, NAÏVE BAYES CLASIFIER, DAN K-NEAREST NEIGHBOR

yang dipersiapkan dan disusun oleh:

Nama : SD. ABDURRAHMAN
Nomor Induk Mahasiswa : 18106050029
Telah diujikan pada : Kamis, 18 Agustus 2022
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Agus Mulyanto, S.Si., M.Kom.
SIGNED

Valid ID: 6304b06c89287



Penguji I
Muhammad Didik Rohmad Wahyudi, S.T.,
MT.
SIGNED

Valid ID: 6304550b63cc9



Penguji II
Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 63048defee5a1



Yogyakarta, 18 Agustus 2022
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6304dad5d0c08



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : SD Abdurrahman

NIM : 18106050029

Judul Skripsi : Analisis Perbandingan Klasifikasi Dokumen Skripsi Menggunakan Support Vector Machine, Naïve Bayes Classifier, dan K-Nearest Neighbor

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 13 Agustus 2022

Pembimbing

Agus Mulyanto, S.Si., M.Kom.

NIP. 19710823 199903 1 003

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : SD Abdurrahman

NIM : 18106050029

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul "**Analisis Perbandingan Klasifikasi Dokumen Skripsi Menggunakan Support Vector Machine, Naïve Bayes Classifier, dan K-Nearest Neighbor**" merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah diajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis dicatu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 13 Agustus 2022

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT. Pemilik alam semesta yang telah melimpahkan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Perbandingan Klasifikasi Dokumen Skripsi Menggunakan Support Vector Machine, Naïve Bayes Clasifier, dan K-Nearest Neighbor”.

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Fakultas Sains dan Teknologi Jurusan Informatika UIN Sunan Kalijaga. Tak dapat dipungkiri pengerjaan skripsi ini sangat memakan tenaga dan membutuhkan keteguhan yang kuat. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

1. Allah SWT. yang telah senantiasa menunjukkan penulis jalan yang benar dan menjaga penulis dari jalan yang sesat.
2. Nabi Muhammad SAW. yang telah membuat penulis tercerahkan dari gelapnya kehidupan.
3. Mendiang ibu penulis, Jamilah yang telah memberikan seluruh ilmu yang ia miliki sampai akhir hayat.
4. Keluarga penulis terutama ayah SD Eryansyah dan adik SD Ahmad Isa dan SY Athiah Saarah yang selalu memberikan dukungan baik yang terucap maupun tidak terucap.
5. Bapak Prof. Dr.Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan

Kalijaga Yogyakarta.

6. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
7. Ibu Ir. Maria Ulfa Siregar, S.Kom., MIT., Ph.D., selaku Ketua Program Studi S1 Teknik Informatika.
8. Bapak Agus Mulyanto, S.Si., M.Kom., selaku Dosen Pembimbing Skripsi dan Akademik yang telah sabar membimbing, mengarahkan, dan memberi nasihat selama penyusunan skripsi.
9. Bapak Didik Rohmat Wahyudi selaku Dosen pengampu mata kuliah *Machine Learning* yang banyak memberi wawasan peneliti dalam penelitian ini.
10. Para dosen yang memverifikasi data pada penelitian ini Bapak Mandahadi Kusuma, M.Eng., Bapak Muhammad Galih Wonoseto, M.T., Bapak Dr. Agung Fatwanto, S.Si., M.Kom., dan Ibu Dwi Otik Kurniawati, M.Eng.
11. Segenap Dosen Fakultas Sains dan Teknologi yang telah mendidik dan memberikan ilmu selama kuliah dan seluruh staf yang selalu sabar melayani segala administrasi selama proses penelitian ini.
12. Teman masa kecil penulis terutama Adhe, Della, Ikram, Nisa, Oja, Yoga dan Yusuf.
13. Teman-teman dan pengasuh pondok pesantren Al-Ashfa bapak Shofiyullah Muzammil.
14. Teman-teman dan pengasuh wisma box.
15. Teman-teman Teknik Informatika angkatan 18.

16. Teman-teman KKN 192 kelompok 105.
17. Teman-teman online server bahas_anime.
18. Dan seluruh teman-teman yang selalu mengiringi penulis dalam proses pengerjaan skripsi Dani, Bayhaqi, Edy, Fadlillah, Fahri, Davin, Ridwan, Galih, May, Ramdani, As'ad, Zhorif, Ihya', dan Fendo.

Semua pihak yang telah membantu dan tidak dapat disebutkan satu persatu.

Semoga tuhan membalas segala pertolongan dan kebaikan yang sudah diberikan kepada saya. Dan pada akhirnya saya menyadari bahwa skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu saya dengan kerendahan hati mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini.

Yogyakarta, Agustus 2022

Penyusun

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

SD Abdurrahman

18106050029

HALAMAN PERSEMBAHAN

*Skripsi ini penulis persembahkan untuk mendiang ibu tercinta
yang dengan tulus selalu memberi penulis ilmu,
pencerahan, serta ridhonya agar penulis
dapat menjadi orang yang sukses di dunia maupun di akhirat.*



HALAMAN MOTTO

“Kita hidup di dunia itu untuk berbahagia,

Jikalau kita sudah bahagia, maka bahagiakanlah orang lain.

Manfaatkan waktumu yang ada sekarang agar kau bisa bahagia selamanya

-Jamilah Baraqbah-



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN COVER.....	i
HALAMAN PENGESAHAN.....	ii
SURAT PERSETUJUAN SKRIPSI.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
HALAMAN PERSEMBAHAN.....	viii
HALAMAN MOTTO.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Batasan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Kebaruan Penelitian.....	5
BAB II.....	6
TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	12
BAB III.....	25
METODE PENELITIAN.....	25
3.1 Kebutuhan Perangkat Sistem.....	26
3.2 Validasi dan Eksplorasi Data.....	27
3.3 Prapengolahan.....	27
3.6 Pelabelan Data.....	28
3.7 Membagi Data Latih dan Data Uji.....	28

3.8 Pembobotan data.....	28
3.9 Implementasi Model	29
BAB IV	30
ANALISIS DAN PERANCANGAN.....	30
4.1 Pengumpulan Data	31
4.2 Validasi dan Analisa Data.....	34
4.3 Pelabelan Data	37
4.4 Pre-processing.....	45
4.5 Pembobotan TF-IDF	67
4.6 Implementasi.....	68
4.7 Evaluasi.....	72
BAB V	77
5.1 Kesimpulan	77
5.2 Saran	78
DAFTAR PUSTAKA	79
LAMPIRAN.....	81
CURICULUM VITAE.....	83

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	26
Gambar 4. 1 Alur Perancangan	30
Gambar 4. 2 Total data.....	31
Gambar 4. 3 Diagram Tema data.....	38
Gambar 4. 4 Confussion Matrix Support Vector Machine	68
Gambar 4. 5 Confussion Matrix Naïve Bayes	69
Gambar 4. 6 Confussion Matrix K-Nearest Neighbor	69
Gambar 4. 7 Diagram perbandingan jumlah data tiap tema	73
Gambar 4. 8 Wordcloud Sistem Cerdas.....	74
Gambar 4. 9 Wordcloud Rekayasa Perangkat Lunak.....	74
Gambar 4. 10 Wordcloud Sistem Komputer dan Jaringan.....	75
Gambar 4. 11 Wordcloud Sistem Informasi	75

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	11
Tabel 2. 2 Ilustrasi Support Vector Machine	18
Tabel 4. 1 Contoh data	31
Tabel 4. 2 Contoh data hasil Validasi.....	35
Tabel 4. 3 Contoh Pelabelan	38
Tabel 4. 4 Contoh penerapan Translation, Case_Folding dan Removing Number and Punctuation	45
Tabel 4. 5 Contoh Tokenisasi.....	55
Tabel 4. 6 Contoh Removing Stopwords	59
Tabel 4. 7 Contoh Stemming	64
Tabel 4. 8 Hasil TF-IDF.....	67
Tabel 4. 9 Contoh Hasil	69



**ANALISIS PERBANDINGAN KLASIFIKASI DOKUMEN SKRIPSI
MENGUNAKAN SUPPORT VECTOR MACHINE, NAÏVE BAYES
CLASIFIER, DAN K-NEAREST NEIGHBOR**

(Studi Kasus : Skripsi Program Studi S1 Teknik Informatika)

SD Abdurrahman

18106050029

INTISARI

UIN Sunan Kalijaga memiliki galeri bahan bacaan berupa skripsi yang terdapat dalam *repository* perpustakaan. Tentunya dengan jumlah yang banyak ini, proses pencarian skripsi yang sesuai akan memakan waktu yang lama dikarenakan harus memeriksanya secara satu persatu. Hal ini dinilai kurang efektif dan efisien sehingga harus ada sebuah sistem yang dapat mempermudah prosesi ini.

Dengan melakukan klasifikasi skripsi yang banyak tersebut dapat terbagi lagi menjadi bagian bagian kecil yang tentunya dapat membuat proses pencarian menjadi lebih efektif dan efisien. Skripsi dengan jumlah banyak tersebut akan terbagi ke dalam empat tema, yaitu: Sistem informasi (SI), Sistem Jaringan Komputer (SJK), Sistem Cerdas (SC), dan Rekayasa Perangkat Lunak (RPL). Namun dalam realisasinya sendiri masih sering didapatkan kekeliruan, dengan memanfaatkan sistem klasifikasi kekeliruan tersebut dapat dikurangi yang akan berdampak baik kedepannya.

Dari hasil penelitian yang dilakukan dengan menggunakan data dari lima tahun terakhir (dari januari 2017 hingga penelitian dibuat) terdapat data dengan jumlah 247, data tersebut terbagi menjadi empat tema, yaitu: Sistem informasi (SI), Sistem Jaringan Komputer (SJK), Sistem Cerdas (SC), dan Rekayasa Perangkat Lunak (RPL), dengan jumlah secara berurutan 13, 25, 130, dan 79 data. Dari pemodelan menggunakan data tersebut *Support Vector Machine* mendapatkan hasil akurasi sebesar 82%, kemudian *Naïve Bayes* sebesar 72%, dan yang terakhir *K-Nearest Neighbors* sebesar 84%. Selain itu dapat disimpulkan juga bahwa SVM lebih peka terhadap data *outliers* dibanding dua algoritma lainnya, dikarenakan nilai *precision* pada SVM dalam menentukan tema data yang paling sedikit yaitu Sistem Jaringan dan Komputer

Kata Kunci: *Support Vector Machine, Naïve Bayes, K-Nearest Neighbors*

**COMPARISON ANALYSIS OF THESIS DOCUMENT CLASSIFICATION
USING SUPPORT VECTOR MACHINE, NAÏVE BAYES CLASIFIER,
AND K-NEAREST NEIGHBOR**

(Case Study : Thesis of Informatics Engineering S1 Study Program)

SD Abdurrahman

18106050029

ABSTRACT

UIN Sunan Kalijaga has a gallery of reading materials in the form of theses contained in the library repository. Of course, with this large number, the process of finding the right thesis will take a long time because you have to check it one by one. This is considered less effective and efficient so there must be a system that can facilitate this process.

By classifying the many theses, it can be further divided into small parts which of course can make the search process more effective and efficient. The large number of theses will be divided into four themes, namely: Information Systems (SI), Computer Network Systems (SJK), Intelligent Systems (SC), and Software Engineering (RPL). However, in the realization itself, errors are still often found, by utilizing the error classification system, it can be reduced which will have a good impact in the future.

From the results of research conducted using data from the last five years (from January 2017 until the research was made) there are 247 data, the data is divided into four themes, named: Information systems (SI), Computer Network Systems (SJK), Intelligent Systems (SC), and Software Engineering (RPL), with a total of 13, 25, 130, and 79 data, respectively. From the modeling using the data, Support Vector Machine gets an accuracy of 82%, then Naïve Bayes is 72%, and finally K-Nearest Neighbors is 84%. In addition, it can also be concluded that SVM is more sensitive to data outliers than the other two algorithms, due to the precision value in SVM in determining the least data theme, namely Network and Computer Systems.

Keywords: *Support Vector Machine, Naïve Bayes, K-Nearest Neighbors*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Program studi teknik informatika adalah salah satu program studi di fakultas sains dan teknologi UIN Sunan Kalijaga, di mana para mahasiswa akan dididik untuk menjadi seseorang yang memiliki kompetensi untuk menyelesaikan permasalahan dalam bidang IT (*Information and Technology*). Sebagai prasyarat kelulusan dari jurusan tersebut, mahasiswa akan diminta untuk menyelesaikan tugas akhir berupa skripsi, "*Nothing new under the sun*" begitulah menurut Charles Darwin, ide ide mengenai pembahasan dalam skripsi sendiri sangat tidak memungkinkan memperoleh bahasan yang benar benar baru, bahasan tersebut tentulah sudah pernah dibahas oleh peneliti lain yang terdahulu, namun hal ini bisa saja menginspirasi untuk melakukan penelitian tentang hal tersebut lebih jauh dan lebih mendetail serta dengan teori dan metode yang baru pula, oleh sebab itu dibutuhkan referensi sebanyak banyaknya untuk mendapatkan hasil penelitian yang lebih maksimal.

Tugas akhir mahasiswa teknik informatika sendiri akan bertambah tiap tahunnya, dari situ dapat kita simpulkan bahwa koleksi tugas akhir informatika ini akan semakin bertambah dan beragam. Koleksi inilah yang nantinya akan menjadi referensi tugas akhir angkatan selanjutnya yang juga akan melakukan penelitiannya dalam pembahasan tertentu yang pastinya akan terkait satu sama lain dengan yang telah ada sebelumnya. Koleksi ini akan sangat berguna jika dapat ditemukan dan dipergunakan dengan baik, ilmu yang sebelumnya akan dapat menginspirasi dan

terus mengalir secara turun temurun ke setiap generasinya. Namun dikarenakan banyaknya skripsi yang ada, menemukan skripsi yang sesuai menjadi hal yang cukup merepotkan dan akan sangat memakan waktu, oleh karena itu sebuah sistem yang dapat menyempitkan ruang pencarian berupa kategorisasi akan sangat membantu

Dengan berkembangnya ilmu pengetahuan setiap waktu menciptakan berbagai macam tema penelitian yang semakin beragam, hal membuat semakin beragamnya skripsi baik dari segi latar belakang, tujuan, metodologi, atau data penelitian. Sehingga kebijakan program studi dalam menyikapi hal ini adalah dengan membagi sekian banyak skripsi tersebut ke dalam empat tema, yaitu: Sistem Informasi (SI), Sistem Jaringan Komputer (SJK), Sistem Cerdas (SC), dan Rekayasa Perangkat Lunak (RPL). Namun dalam realisasinya sendiri masih sering didapatkan kekeliruan dalam pengelompokannya, yang akan menyebabkan sulitnya mahasiswa untuk menemukan skripsi atau referensi yang sesuai dengan bidang mereka, dengan memanfaatkan sistem klasifikasi kekeliruan tersebut dapat dikurangi yang akan berdampak baik kedepannya.

Sistem klasifikasi di sini memegang peranan penting, karena dinilai perlu mengingat sulitnya mendapatkan informasi mengenai hal tersebut. Sistem ini diharapkan dapat mempermudah hal tersebut dan akan lebih menghemat waktu dan tenaga, karena untuk membaca dan menemukan skripsi tersebut mahasiswa diharuskan untuk datang ke perpustakaan Universitas dan mencarinya satu persatu.

Perpustakaan di Universitas Islam Sunan Kalijaga memang sudah dilengkapi dengan *repository* atau galeri skripsi yang sangat banyak, karena banyaknya koleksi

yang dimiliki inilah yang membuat skripsi sulit ditemukan, oleh karena itu sistem klasifikasi dinilai perlu untuk membagi lagi skripsi yang banyak tersebut ke dalam beberapa golongan guna mempersingkat proses pencarian.

Menurut Clara Bridge (2011), dikarenakan banyaknya informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, text mining diyakini memiliki potensi nilai komersial tinggi. Hal yang lebih menyulitkan dalam analisis adalah bahwa sekitar 80% sampai 85% bentuk informasi tersebut dalam format tidak terstruktur (*unstructured data*).

Text Mining atau dikenal juga dengan analisa teks adalah suatu proses pengambilan intisari dari dokumen teks sehingga didapatkan hasil yang berguna untuk tujuan tertentu. *Text mining* dapat diartikan sebagai suatu proses dalam mendapatkan informasi lewat penggalian mendalam terhadap sejumlah data yang tidak terstruktur untuk mendapatkan sejumlah informasi dan pola yang nantinya akan diolah lebih dalam lagi. Salah satu implementasi dari *text mining* yaitu klasifikasi. Klasifikasi di sini dimaksudkan untuk memudahkan pengelompokan data yang besar lewat pola tersebut, tanpa harus memisahkannya secara manual dengan memeriksanya satu per satu.

Sistem klasifikasi sendiri merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua proses yang dilakukan yaitu dengan membangun model untuk disimpan sebagai memori dan menggunakan model tersebut untuk melakukan pengenalan atau klasifikasi atau prediksi pada suatu data lain supaya diketahui di kelas mana objek data tersebut dimasukkan berdasarkan model yang

telah disimpan dalam memori (Prasetyo, 2012).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini adalah:

1. Bagaimana cara untuk mengelompokkan skripsi dengan tepat dan cepat untuk menunjang proses pencarian yang dilakukan mahasiswa di *website repository* perpustakaan UIN Sunan Kalijaga?
2. Apa Algoritma terbaik untuk mengklasifikasikan skripsi bidang Teknik Informatika di UIN Sunan Kalijaga?

1.3 Tujuan

Penelitian ini bertujuan untuk:

1. Mempermudah mahasiswa Informatika untuk menemukan skripsi berupa skripsi yang relevan lewat pengklasifikasian tema.
2. Mengetahui akurasi dari ketiga algoritma yang akan digunakan dalam pengklasifikasian tema skripsi, yaitu: *Naïve Bayes Classifier*, *Support Vector Machine*, dan *K-Nearest Neighbors*.

1.4 Batasan Penelitian

Dalam penelitian ini akan diterapkan beberapa batasan untuk memaksimalkan penelitian agar tidak melebar dan mengurangi keakuratan penelitian. Batasan yang digunakan dalam penelitian ini adalah Penelitian ini hanya menggunakan data yang didapat dari <https://digilib.uin-suka.ac.id/> dengan tema Teknik Informatika dan hanya berupa abstrak dan judul skripsi tugas akhir dari mahasiswa S1 saja terhitung

sejak tanggal 1 Januari 2017 sampai penelitian ini dimulai (6 Februari, 2022).

Dalam penelitian ini, peneliti hanya menggunakan tiga algoritma, yaitu: *Naïve Bayes Classifier*, *Support Vector Machine*, dan *K-Nearest Neighbors*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan akan memiliki manfaat sebagai berikut:

- Bagi sivitas akademika yang berkaitan dengan Informatika, dapat mempermudah dalam memilih skripsi terkait tugas akhir yang sesuai dengan tema yang sudah diklasifikasikan.
- Bagi UIN Sunan Kalijaga, penelitian ini dapat menjadi acuan dalam pengembangan sistem informasi tugas akhir yang akan digunakan di universitas.
- Bagi penulis, dapat memberikan pengetahuan dan wawasan yang baru.
- Bagi pembaca, dengan adanya penelitian ini dapat digunakan sebagai sumber pembelajaran untuk penelitian selanjutnya dan menambah wawasan baru

1.6 Kebaruan Penelitian

Sudah sangat marak ditemukan penelitian terkait tentang klasifikasi bacaan, namun hal yang menurut saya baru adalah sistem bacaan yang sesuai dan terdapat dalam koleksi tugas akhir di Perpustakaan UIN Sunan Kalijaga.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian dan eksperimen yang sudah dilakukan, disimpulkan bahwa pada penelitian ini data bacaan tugas akhir bidang informatika didapatkan dari *website* perpustakaan UIN Sunan Kalijaga pada tanggal 6 Februari, 2022. Data awal yang berjumlah 743 kemudian dilakukan penyaringan dengan beberapa persyaratan seperti tahun keluar penelitian, jenis penelitian, dan sebagainya, sehingga hanya menyisakan data sebesar 247 data saja.

Setelah dilakukan *labeling*, data terbagi menjadi empat tema, yaitu: Sistem informasi (SI), Sistem Jaringan Komputer (SJK), Sistem Cerdas (SC), dan Rekayasa Perangkat Lunak (RPL), dengan jumlah secara berurutan 13, 25, 130, dan 79 data. Dari hal ini dapat disimpulkan bahwa dalam kurun waktu lima tahun terakhir mahasiswa cenderung memilih tema Sistem Cerdas dan tema Sistem Jaringan dan Komputer merupakan tema yang paling sedikit. Kemudian data yang sudah diterapkan *pre-processing*, kemudian data akan diinputkan ke dalam model untuk dilatih yang nantinya akan diukur akurasi.

Support Vector Machine mendapatkan hasil akurasi sebesar 82%, kemudian *Naïve Bayes* sebesar 72%, dan yang terakhir *K-Nearest Neighbors* sebesar 84%. Akurasi ini juga masih belum bisa dipastikan, dikarenakan jumlah data yang tidak setara antara tema satu dan lainnya. Dan juga dapat disimpulkan dari penelitian ini KNN lah yang mendapat nilai tertinggi dalam mengklasifikasi abstrak pada skripsi untuk menentukan temanya, selain itu dapat disimpulkan juga bahwa SVM lebih

peka terhadap data *outliers* dibanding dua algoritma lainnya, dikarenakan nilai *precision* pada SVM dalam menentukan tema data yang paling sedikit yaitu Sistem Jaringan dan Komputer.

5.2 Saran

Pada penelitian ini tentunya tidak terlepas dari kelemahan dan kekurangan, oleh karena itu peneliti menyarankan adanya perbaikan-perbaikan dan langkah langkah yang dapat dijadikan masukan untuk peneliti selanjutnya, adapun sebagai berikut :

1. Data harus lebih banyak dan lebih merata serta beragam, hal ini tentu dapat meningkatkan kualitas data latih yang akan dimasukkan ke dalam model.
2. Menggunakan algoritma lain khususnya algoritma yang cukup handal untuk menangani *outliers* dan *overfitting* sehingga akurasi yang didapat akan lebih baik pula.
3. Mencoba melihat masalah dari perspektif berbeda, mungkin dengan pembagian kelas yang berbeda, misal dengan melihat metode, objek dsb.
4. Mencoba menghubungkannya dengan clustering untuk mendapatkan *knowledge* yang lebih beragam.
5. Melakukan *labeling* dengan tenaga ahli.

DAFTAR PUSTAKA

- Adityo P. W. , (2016) Sistem Klasifikasi Kinerja Satpam Menggunakan Metode Naïve Bayes Classifier
- Andreas K. dan Michael H., (2019). User Of The World, Unite! The Challenges and Opportunities Of Social Media
- Arthur S., (1959). Some Studies in Machine Learning Using the Game of Checkers.
- Ayu F., (2018). Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier
- C. Bridge. 2011. Unstructured Data and the 80 Percent Rule.
- Dea H. K., (2017). Implementasi Algoritma Multinomial Naive Bayes Classifier.
- Diah B. A., (2021). Pengembangan Sistem Klasifikasi Kepustakaan dengan Dewey Decimal Classification (DDC).
- Ghazanfar, M. A., & Prügel-Bennett, A., (2010). Leveraging Clustering Approaches To Solve The Gray-Sheep Users Problem In Recommender Systems.
- Larose .(2005). Discovering Knowledge in Data : an Introduction to Data Mining.
- Ni Luh Gede P. S. , (2017) Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil.
- Nur S. ,(2018). Sistem Rekomendasi Pemilihan Smartphone Snapdragon 636 Menggunakan Metode Simple Multi Attribute Rating Technique.
- Parveen & Pandey, (2017). Sentiment Analysis on Twitter Data-set Using Naïve Bayes Algorithm
- Prasetyo, (2012). Data Mining Konsep dan Aplikasi Menggunakan Matlab.

Ronen F. & James S. (2007). The Text Mining Handbook: Advanced Approach in
Analyzing Unstructured Data.

Santosa, (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.

Thomas C. ,(2005). Elements of Information Theory.

Vapnik V., (1995). Support-Vector Networks.

