

**FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN AN-
ALISIS LIRIK LAGU**

Skripsi



oleh:

ELDURDUZ MAHARDHIKA PALAR DEWANTARA

71160115

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

2022

**FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN AN-
ALISIS LIRIK LAGU**

Skripsi



Diajukan kepada Fakultas Teknologi Informasi Program Studi In-
formatika Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

oleh:

ELDURDUZ MAHARDHIKA PALAR DEWANTARA

71160115

DUTA WACANA
PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2022

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Elduarduz Mahardhika Palar Dewantara
NIM : 71160115
Program studi : Informatika
Fakultas : Fakultas Teknik Informasi
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN ANALISIS LIRIK LAGU

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 12 April 2020

Yang menyatakan



(Elduarduz Mahardhika Palar Dewantara)

71160115

HALAMAN PENGESAHAN

FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN ANALISIS LIRIK LAGU

Oleh: ELDUARDUZ MAHARDHIKA PALAR D / 71160115

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 21 Juni 2022

Yogyakarta, 9 Juli 2022
Mengesahkan,

Dewan Penguji:

1. Lukas Chrisantyo, S.Kom., M.Eng.
2. Antonius Rachmat C., S.Kom., M.Cs.
3. Yuan Lukito, S.Kom., M.Cs.
4. Gloria Virginia, S.Kom., MAI, Ph.D.



Dekan

(Restyandito, S.Kom., MSIS., Ph.D.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN ANALISIS LIRIK LAGU

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 9 Juli 2022



ELDUARDUZ MAHARDHIKA
PALAR D
71160115

HALAMAN PERSETUJUAN

Judul Skripsi : FITUR KLASIFIKASI LAGU LITURGI KATOLIK
BERDASARKAN ANALISIS LIRIK LAGU

Nama Mahasiswa : Elduarduz Mahardhika Palar Dewantara

N I M : 71160115

Matakuliah : Skripsi

Kode : TI0341

Semester : Genap

Tahun Akademik : 2020/2021

Telah diperiksa dan disetujui
Di Yogyakarta
Pada tanggal 31/05/2021,

Dosen Pembimbing I

Dosen Pembimbing II



Lukas Chrisantyo A A.,
S,Kom., M. Eng.



Antonius Rachmat C.,
S.Kom.,M.Cs.

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Skripsi dengan judul FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN ANALISIS LIRIK LAGU ini telah selesai disusun.

Penulis memperoleh banyak bantuan dari kerja sama baik secara moral maupun spiritual dalam penulisan Skripsi ini, untuk itu tak lupa penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Tuhan yang maha kasih, yang telah memberkati setiap perjalanan dan mengerjakan keseluruhan skripsi
2. Orang tua yang selama ini telah sabar membimbing dan mendoakan penulis tanpa kenal untuk selama-lamanya,
3. Restyandito, S.Kom, MSIS., Ph.D selaku Dekan FTI juga selaku dosen wali yang telah memberikan masukan dan dorongan untuk menempuh perkuliahan secara keseluruhan.
4. Gloria Virginia, S.Kom., MAI, Ph.D. Kaprodi selaku Kaprodi Informatika. Yang telah memberikan semangat secara keseluruhan terhadap mahasiswa.
5. Lukas Chrisantyo A A., S,Kom., M. Eng, selaku Dosen Pembimbing 1, yang telah memberikan ilmunya dan dengan penuh kesabaran membimbing penulis.
6. Antonius Rachmat C., S.Kom.,M.Cs., selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan kesabaran dalam membimbing penulis.
7. Keluarga tercinta: yang selalu memberikan semangat, dukungan, dan nasehat serta doa yang memberikan motivasi untuk penyelesaian penelitian ini.
8. Seluruh staff pengajar dan karyawan Jurusan Informatika UKDW.
9. Lain-lain yang telah mendukung moral, spiritual, dan dana untuk belajar selama ini.

Laporan proposal/skripsi ini tentunya tidak lepas dari segala kekurangan dan kelemahan, untuk itu segala kritikan dan saran yang bersifat membangun guna

kesempurnaan skripsi ini sangat diharapkan. Semoga proposal/skripsi ini dapat bermanfaat bagi pembaca semua dan lebih khusus lagi bagi pengembangan ilmu komputer dan teknologi informasi.

Yogyakarta, 02/06/2022



Elduarduz M P Dewantara



DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSETUJUAN.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
ABSTRACT.....	xvi
1. BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian	4
1.6.1 Diagram Blok Sistem	4
1.6.2 Sistem Aplikasi	5
1.7 Sistematika Penulisan.....	6
2. BAB II LANDASAN TEORI.....	7
2.1 Tinjauan Pustaka	7
2.2 Dasar Teori	8
2.2.1 <i>Machine Learning</i>	8
2.2.2 Teknik Klasifikasi	9

2.2.3	<i>Dataset Collection</i>	11
2.2.4	<i>Pre-processing</i>	11
2.2.5	<i>Features Selection, Extraction/Creation</i>	13
2.2.6	<i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	14
2.2.7	<i>KNN (K-Nearest Neighbors)</i>	15
2.2.8	<i>(2-8)Precision & Recall</i>	17
2.2.9	<i>F-Measure</i>	18
3.	BAB III METODE PENELITIAN	20
3.1	Analisis Kebutuhan	20
3.1.1	Kebutuhan Fungsional & Non-Fungsional	20
3.1.2	Kebutuhan Perangkat Keras	21
3.1.3	Kebutuhan Perangkat Lunak	21
3.1.4	Kebutuhan Teknologi dan <i>Library</i>	21
3.2	Profil Dataset	21
3.2.1	Data Lagu	21
3.2.2	Data Kategori Lagu	22
3.3	Blok Diagram	23
3.3.1	Rancangan Proses	23
3.3.2	Rancangan Sistem Aplikasi	29
3.3.3	Use Case Diagram	31
3.4	Metode Pengumpulan Data	31
3.5	Metode Teks Pre-Processing	32
3.6	Rancangan Basis Data	34
3.7	Mock Up	35
3.8	Metode Pengujian Sistem	38

4. BAB IV HASIL PENELITIAN DAN PEMBAHASAN	40
4.1 Implementasi Sistem	40
4.1.1 Implementasi Antarmuka	40
4.1.2 Teks Pre-Processing	42
4.1.3 Teknik TF-IDF	44
4.1.4 Klasifikasi	45
4.2 Pengujian	46
4.2.1 Hasil Pengujian	46
4.2.2 Hasil Pengujian	46
4.3 Evaluasi Sistem	72
4.3.1 Pengujian sistem fitur klasifikasi lirik lagu.....	72
4.3.2 Label/ Kategori Lagu	74
4.3.3 Text Pre-Processing	74
4.3.4 Hasil Pengujian Nilai K	77
4.3.5 K-Value (Error Rate).....	79
5. BAB V KESIMPULAN DAN SARAN.....	81
5.1 Kesimpulan.....	81
5.2 Saran.....	81
Daftar Pustaka	83
6. LAMPIRAN.....	86
LAMPIRAN 1 – Hasil Precision, Recal, F1-Score.....	86
LAMPIRAN 2 – <i>Source Code</i> Pre-processing data lirik	88
LAMPIRAN 3 – <i>Source Code</i> Klasifikasi Multi-Label KNN.....	90
LAMPIRAN 4 – Kartu Konsultasi.....	95

DAFTAR GAMBAR

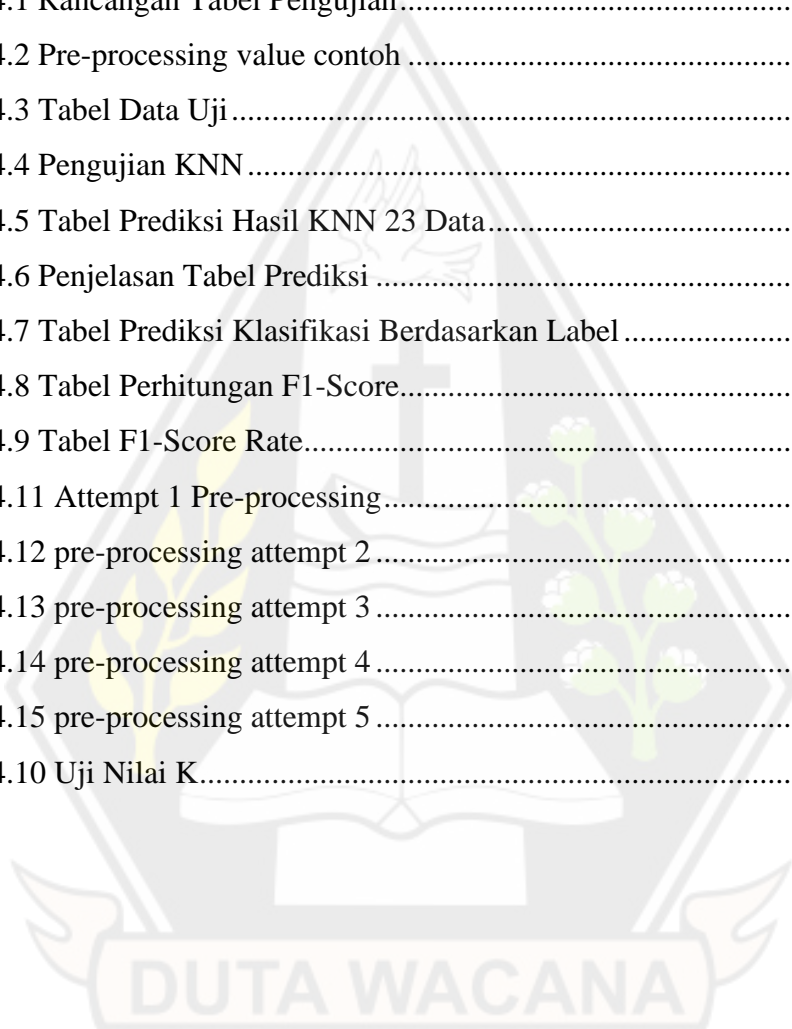
Gambar 1.1 Block Diagram Penelitian	4
Gambar 1.2 Blok Diagram Sistem	5
Gambar 2.1 Hierarchy of Artificial Intelligence (Chamunyonga, Edwards, Caldwell, & Burberry, 2020).....	9
Gambar 2.2 Data Preparation Sample (Gracia, Luengo, & Herrera, 2014).....	12
Gambar 2.3 Data Reduction Sample (Gracia, Luengo, & Herrera, 2014).....	13
Gambar 3.1 Blok Diagram Penelitian	23
Gambar 3.2 Teknik Pengumpulan Data	25
Gambar 3.3 Pre-Processing Alur.....	26
Gambar 3.4 TF-IDF Process	27
Gambar 3.5 KNN Proses.....	28
Gambar 3.6 Alur Evaluasi.....	29
Gambar 3.7 Blok Diagram Sistem	30
Gambar 3.8 Use Case Diagram.....	31
Gambar 3.9 Website POMUKA Tampilan Depan.....	32
Gambar 3.10 Rancangan Basis Data.....	35
Gambar 3.11 Mock Up Layout Klasifikasi	36
Gambar 3.12 Proses Klasifikasi	37
Gambar 4.1 Antarmuka Klasifikasi.....	41
Gambar 4.2 Layout klasifikasi	42
Gambar 4.3 Sebelum Casefolding.....	42
Gambar 4.4 Setelah Casefolding.....	42
Gambar 4.5 Sebelum Punctuation.....	43
Gambar 4.6 Sesudah Punctuation	43
Gambar 4.7 Hasil Tokenizing	43
Gambar 4.8 Distribution Frequency (Belum Sempurna)	43
Gambar 4.9 Sebelum Stopword Removal.....	43
Gambar 4.10 Setelah Stopword Removal	43
Gambar 4.11 Sebelum Stemming	44
Gambar 4.12 Sesudah Stemming	44

Gambar 4.13 Hasil TF-IDF	45
Gambar 4.14 Hasil Klasifikasi	45
Gambar 4.15 Form Input Data	47
Gambar 4.16 Contoh Pengambilan Data Klasifikasi	53
Gambar 4.17 Layout Input Tambah Lagu	72
Gambar 4.18 Layout Klasifikasi	73
Gambar 4.19 Proses Klasifikasi	73
Gambar 4.20 K value elbow method	80



DAFTAR TABEL

Tabel 3.1 Kebutuhan Fungsional	20
Tabel 3.2 Kebutuhan Non-Fungsional	20
Tabel 3.3 Data Berdasarkan Kategori	22
Tabel 3.4 Jenis Pre-Processing Sistem.....	38
Tabel 4.1 Rancangan Tabel Pengujian.....	46
Tabel 4.2 Pre-processing value contoh	48
Tabel 4.3 Tabel Data Uji.....	53
Tabel 4.4 Pengujian KNN	58
Tabel 4.5 Tabel Prediksi Hasil KNN 23 Data.....	68
Tabel 4.6 Penjelasan Tabel Prediksi	69
Tabel 4.7 Tabel Prediksi Klasifikasi Berdasarkan Label	69
Tabel 4.8 Tabel Perhitungan F1-Score.....	70
Tabel 4.9 Tabel F1-Score Rate.....	71
Tabel 4.11 Attempt 1 Pre-processing	74
Tabel 4.12 pre-processing attempt 2	75
Tabel 4.13 pre-processing attempt 3	75
Tabel 4.14 pre-processing attempt 4	75
Tabel 4.15 pre-processing attempt 5	76
Tabel 4.10 Uji Nilai K.....	78



INTISARI

FITUR KLASIFIKASI LAGU LITURGI KATOLIK BERDASARKAN ANALISIS LIRIK LAGU

Indonesia adalah salah satu negara yang memiliki ragam adat yang bisa terbilang banyak begitupun ragam agama yang ada. Salah satunya yaitu adalah agama Kristiani. Pada setiap minggunya agama Katolik mengadakan ibadah yang disebut dengan Ekaristik. Dimana, pada saat beribadah dan menyanyikan lagu tertentu pastinya akan menjadi kesulitan untuk mencari lagu dengan cepat. Maka, dibuatkanlah Website POMUKA (Pencarian Online Musik Katolik) yang berguna untuk membantu pencarian lagu. Namun, pada jaman sekarang ini banyaknya perkembangan musik katolik yang tidak dapat dihambat, membuat setiap pencinta musik lagu katolik tidak mengerti terhadap jenis lagu yang didengarkan.

Berdasarkan penjelasan diatas, penelitian ini bertujuan untuk membuat sebuah fitur yang dapat melakukan klasifikasi pada lagu disebuah aplikasi website yang sudah ada bernama POMUKA. Pengklasifikasian akan dilakukan dengan menggunakan K Nearest Neighbors (KNN). Pengklasifikasian KNN ini dapat terbilang berbeda, karena dapat mampu menemukan label atau kategori yang diinginkan lebih dari satu sehingga pengguna bisa memilah kategori mana saja yang cocok pada lirik lagu yang ingin diketahui. Agar pengklasifikasian berjalan dengan baik, ada beberapa penggunaan metode pendukung seperti *pre-processing* data untuk menormalisasikan data lagu yang ada. Juga, *Term Frequency – Inverse Document Frequent* (TF-IDF) yang akan digunakan untuk memboboti sebuah teks yang akan proses nantinya. Namun, untuk meningkatkan performa klasifikasi pada biasanya, pada penelitian ini akan membandingkan penggunaan *pre-processing* yang ada. Dimana, ada beberapa perbedaan di dalam *pre-processing* yang diterapkan yaitu membandingkan penggunaan atau tidaknya terhadap case folding, punctuation, tokenizing, stopword removal, dan stemming sastrawi.

Pada penelitian ini pengklasifikasian dengan menggunakan K Nearest Neighbors membutuhkan pengukuran akurasi yang berguna untuk menilai tingkat ketepatan. Maka, penggunaan F-Measure akan berguna untuk penelitian ini.

Kata kunci: machine learning, K Nearest Neighbors, term frequency inverse document frequency, manhattan distance, pre-processing.



ABSTRACT

Indonesia is one of the countries that has a variety of customs that can be said to be many as well as the variety of religions that exist. One of them is Christianity. Every week the Catholic religion holds a service called the Eucharist. Where, when worshiping and singing certain songs, it will certainly be difficult to find songs quickly. So, the POMUKA (Catholic Music Online Search) Website was created which is useful to help search songs.

However, nowadays there are many developments of Catholic music that cannot be inhibited, making every Catholic music lover not understand the type of song that is listened to. Based on the explanation above, this study aims to create a feature that can classify songs in an existing website application called POMUKA. Classification will be carried out using K Nearest Neighbors (KNN). This KNN classification can be quite different, because it can be able to find more than one desired label or category so that users can sort out which categories are suitable for the song lyrics they want to know. In order for the classification to run well, there are several uses of supporting methods such as pre-processing data to normalize the existing song data. Also, Term Frequency – Inverse Document Frequent (TF-IDF) which will be used to weight a text that will be processed later. However, to improve classification performance in general, this study will compare the use of existing pre-processing.

Where, there are several differences in the pre-processing applied, namely comparing the use or not of case folding, punctuation, tokenizing, stop-word removal, and literary stemming. In this study, classification using K Nearest Neighbors requires an accuracy measurement that is useful for assessing the level of accuracy. So, the use of F-Measure will be useful for this research.

BABI

PENDAHULUAN

1.1 Latar Belakang

Fitur pengkategorian merupakan salah satu bentuk penyederhanaan yang mengelompokkan berbagai informasi sehingga mudah untuk melakukan suatu pencarian (Sulartopo, 2015). Saat melakukan pencarian pada data yang telah dikategorikan tentunya akan mempermudah pencarian jenis data yang akan ditemukan nantinya. Website Pomuka (Pencarian Online Musik Katolik) adalah salah satu website pencarian lagu katolik yang merupakan sebuah project dari Bp. Lukas Chrisantyo yang dimana website tersebut juga memiliki fitur perekaman data lagu katolik dan juga fitur pencarian lagu berdasarkan frase/kalimat Kitab Suci yang dimasukkan. Namun, ada kendala yang ada saat ini yaitu ketika seorang ingin mengklasifikasikan kategori lagu berdasarkan lirik lagu saja dan tidak menggunakan judul lagu yang diketahui. Sehingga hanya dengan mengetahui lirik lagu saja, pengguna bisa melakukan pengkategorian lagu sesuai yang diinginkan tentunya dengan menggunakan lirik yang diingat.

Penggunaan metode klasifikasi dalam sebuah fitur pengkategorian sudah dilakukan oleh beberapa orang pada penelitian sebelumnya dan akhirnya berfungsi dengan baik sehingga bisa mengkategorikan jenis lagu dan menemukan lagu apa yang ingin dicari dengan mudah. Sebagai contoh penggunaan metode *naïve bayes* dengan pembobotan TF-IDF pada pengklasifikasian lagu daerah dengan menggunakan lirik oleh (Saputro, Aristian, & Tyas, 2017) menjadi salah satu penelitian yang patut dilakukan. Adapun pengklasifikasian emosi pada lagu dengan lirik berbahasa Indonesia oleh (Armianti, Indriati, & Adinugroho, 2019) dimana metode yang digunakan yaitu *k-Nearest Neighbors* dengan metode pembobotan WIDF. Namun, hasil dari akurasi yang didapat kurang memuaskan dengan persenan 66% pada tahap akhir. Oleh karena itu, pembobotan yang dilakukan saat ini dengan menggunakan TF-IDF semoga akan lebih baik dibandingkan pembobotan yang dilakukan pada penelitian sebelumnya. Namun pada penelitian kali ini,

penggunaan *k-Nearest Neighbors* akan dipakai untuk pengujian yang akan dilakukan.

Untuk melakukan pencarian dan pengelompokan suatu lagu dengan cepat dan akurat, sistem pencarian dalam suatu website harus memiliki metode agar bisa berjalan dengan baik. Penerapan metode pemanfaatan dari lirik lagu yang sudah diketahui sebelumnya sangatlah dibutuhkan sehingga bisa dipaparkan berdasarkan lagu yang akan dikategorikan. Pencarian lagu dalam sebuah website sebenarnya sudah sangat sering dilakukan. Namun, pencarian lirik lagu dan mengelompokkan lagu berdasarkan beberapa hal memang masih terbilang tidak banyak sehingga beberapa orang yang hanya mengetahui lirik saja, akan lebih terbantu ketika lupa dengan judul lagu yang ada. Biasanya seseorang akan mencari lagu dengan memasukkan lirik lagu yang dicarinya dan akan menemukan lagu yang dicari, tetapi tidak berdasarkan pengkategorian genre atau jenis lagu dan sejenisnya tergantung pengkategorian yang diinginkan.

Fitur pengklasifikasian ini secara singkat adalah fitur dengan jenis klasifikasi kategori lagu, dimana sebuah lagu yang sudah ada akan dipilah sesuai inputan yang diberikan dan mendapatkan jenis-jenis lagu yang cocok menurut inputan yang diberikan. Sebagai contoh, pada saat pengguna ingin mencari kategori lagu yang ada dengan informasi yaitu lirik lagu yang diketahui, pengguna akan melakukan input data kemudian *apply* dan fitur ini akan mengklasifikasikan jenis lagu yang ada dalam database sesuatu program yang sudah dibuat dan akan muncul pada tampilan yang tersedia berupa list kategori/jenis lagu sesuai lirik lagu yang diinputkan oleh pengguna.

Maka dari itu, penelitian yang bertujuan untuk merancang sebuah sistem pengkategorian jenis lagu dengan lirik lagu dengan baik. Akhirnya akan dibuat dengan menggunakan metode *KNN* yang berguna untuk melakukan pengklasifikasian dan juga TF-IDF sebagai metode pembobotan teks lirik lagu. Metode *KNN* ini sangatlah penting walaupun algoritma ini bersifat *lazy learning* namun, keakuratannya akan lebih meningkat secara teknis. Website yang nantinya akan

ditambahkan fitur bertujuan untuk membantu pengguna mencari kategori lagu yang diinginkan berdasarkan lirik lagu yang diinputkan.

1.2 Rumusan Masalah

Seperti yang sudah dijelaskan mengenai latar belakang diatas, maka penulis merumuskan rumusan masalah sebagai berikut:

Sejauh mana tingkat keberhasilan pengklasifikasian lagu berdasarkan analisis lirik dengan menggunakan *k-nearest neighbors* sebagai metode pengklasifikasian jenis lagu?

1.3 Batasan Masalah

Berikut merupakan batasan masalah yang ditemukan dalam penelitian:

1. Lirik lagu diambil dari buku Puji Syukur dan Madah Bakti yang sudah terhimpun di web Pomuka.
2. Lirik lagu yang digunakan bukanlah merupakan lirik lagu berbahasa diluar bahasa Indonesia.
3. Teknik *pre-processing* yang akan digunakan akan disesuaikan berdasarkan kebutuhan, dan tidak semuanya.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu untuk mengetahui seberapa jauh tingkat keberhasilan yang diraih dari pengklasifikasian lagu berdasarkan lirik lagu dengan *pre-processing* yang baik dan metode *k-nearest neighbors* sebagai suatu metode *classification* yang akan digunakan nantinya.

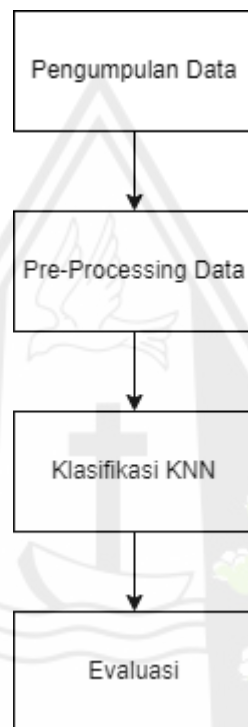
1.5 Manfaat Penelitian

Dengan adanya penelitian ini adapun manfaat yang bisa didapat adalah membantu pengguna untuk menemukan kategori yang cocok dari lagu dengan menggunakan lirik lagu sehingga bisa mencari kategori lagu yang tepat nantinya.

1.6 Metodologi Penelitian

1.6.1 Diagram Blok Sistem

Diagram blok sistem yang akan dibuat seperti pada Gambar 1.1 Block Diagram Penelitian.



Gambar 1.1 Block Diagram Penelitian

Keterangan mengenai Gambar 1.1 Block Diagram Penelitian sebagai berikut:

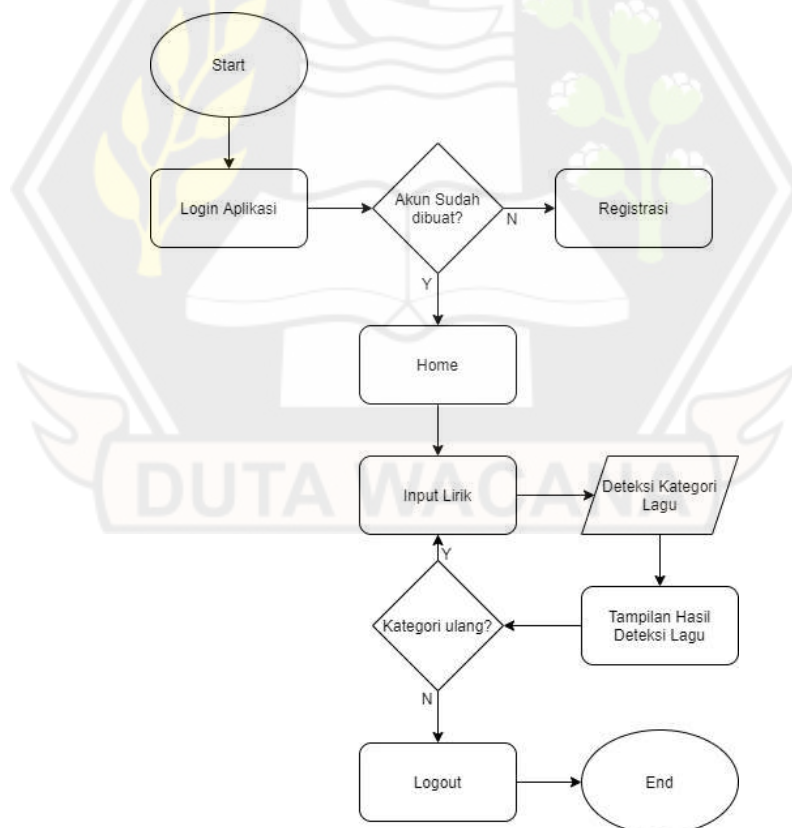
1. Pengumpulan data yaitu menentukan data yang akan digunakan nantinya untuk dijadikan subjek data processing sehingga bisa dikategorikan.
2. *Preprocessing* data yaitu melakukan prosesing data dimana data yang sudah ada akan diseragamkan dalam suatu format tertentu sehingga kumpulan data yang ada tidak dalam format yang berbeda yang dimana pada akhirnya akan mengoptimalkan akurasi dalam melakukan klasifikasi. Pada tahap ini, tentunya akan ada penerapan feature extraction dan feature selection.
3. *K-Nearest Neighbors* yaitu suatu metode klasifikasi yang akan dipakai pada penelitian ini. Dengan menggunakan metode ini, pengecekan disetiap data pada teks lirik akan lebih lambat, namun menjanjikan karena disetiap sisi

akan dibedakan satu persatu berdasarkan *lazy learning*. Sehingga mencapai titik maksimal yaitu nilai k .

4. Evaluasi adalah tahap untuk melakukan validasi dan testing dari segi keakuratan fitur yaitu pengecekan tingkat akurasi terhadap fitur yang sudah dibuat, apakah sudah mencerminkan kesamaan data dari fitur yang diinginkan atau tidak. Semakin akurat data yang dicari tentunya semakin baik mencerminkan metode yang digunakan pada fitur tersebut. Pada tahap ini, hasil dari metode klasifikasi akan dilakukan test ke akuratnya dengan menggunakan pengukuran kinerja klasifikasi yaitu F-Measure sehingga bisa mengetahui tingkat keakuratan fitur yang akan dibuat berdasarkan nilai presisi dan recall dari dataset klasifikasi yang ada.

1.6.2 Sistem Aplikasi

Rancangan sistem aplikasi yang akan dibuat seperti pada Gambar 1.2 Blok Diagram Sistem.



Gambar 1.2 Blok Diagram Sistem

1.7 Sistematika Penulisan

Laporan tugas akhir yang disusun akan terdiri dari lima bab sehingga bisa dibagi menjadi Pendahuluan, Tinjauan Pustaka dan Landasan Teori, Metodologi Penelitian, Hasil dan Pembahasan, dan yang terakhir Kesimpulan dan Saran.

1. Bab I (Pendahuluan)

Bagian ini akan menjelaskan mengenai gambaran secara umum dari penelitian. Berikut uraian dari bab I yakni latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian sistematika penulisan.

2. Bab II (Tinjauan Pustaka dan Landasan Teori)

Bagian ini terdiri dari dua bagian utama yakni Tinjauan Pustaka dan Landasan Teori. Pada Tinjauan Pustaka akan berisi tentang artikel jurnal yang telah dipublikasikan dan relevan dengan topik penelitian. Pada bagian Landasan Teori akan berisi tentang teori – teori, rumus dan definisi yang berkaitan topik dan akan digunakan dalam penelitian.

3. Bab III (Metodologi Penelitian)

Bagian ini akan membahas mengenai tahapan – tahapan penelitian dan perancangan sistem seperti diagram penelitian, spesifikasi pembuatan sistem dan metode penelitian.

4. Bab IV (Hasil dan Pembahasan)

Bagian ini akan berisi hasil dari implementasi pengembangan aplikasi dan evaluasi rancangan antarmuka penelitian

5. Bab V (Kesimpulan dan Saran)

Bagian ini akan berisi kesimpulan yang didapat dari hasil penelitian yang telah dilakukan dan saran yang akan diberikan oleh peneliti untuk pengembangan penelitian selanjutnya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi sistem dan analisis sistem, penggunaan metode KNN untuk pengklasifikasian kategori lagi berdasarkan lirik lagu dimana, lagu yang dipakai yaitu lagu liturgi katolik terfokus kepada Madah Bakti dan Puji Syukur yang sebagaimana data ini sangat mempengaruhi hasil dari penelitian sehingga menemukan hasil average akurasi 0.4344 dan hasil average f1score 0.3530 yang dimana ini adalah hasil tertinggi yang didapat dengan menggunakan pre-processing stemming sastrawi menggunakan nilai $k=3$.

Maka, bisa disimpulkan bahwa penggunaan metode KNN untuk pengkategorian lagu liturgi katolik ini masih dibawah rata-rata dan kurang begitu baik untuk lagu liturgi katolik yang berfokus pada Madah Bakti dan Puji Syukur dengan hasil rata-rata akurasi 43,44%.

5.2 Saran

Penelitian ini tentunya dapat dikembangkan menjadi lebih baik lagi. Berikut adalah beberapa saran yang dapat disampaikan untuk penelitian selanjutnya:

1. Menambah jumlah data testing yang dipakai, semisal 20%-30% data dari data uji sebagai data testing yang bertujuan untuk dapat melihat secara lebih luas lagi tingkat pengklasifikasian.
2. Menggunakan k-fold sebagai sistem evaluasi untuk dapat mencari nilai atau tingkat akurasi yang tepat, karena k-fold melakukan pengujian terhadap keseluruhan data yang ada sehingga dapat melihat atau mengevaluasi dengan dari cakupan yang luas. Dan dapat membandingkan dengan evaluasi f-measure.
3. Bisa menggunakan metode yang lain untuk melakukan penelitian dengan lirik lagu Madah Bakti dan Puji Syukur karena hasil evaluasi dengan KNN kurang baik. Metode yang mungkin bisa dicoba adalah Naïve Bayes dimana, metode tersebut menggunakan mekanisme probabilitas.

4. Data yang digunakan harus diolah dengan baik sebagai contoh, pembobotan nilai dari setiap lirik yang ada harus setara. Pembobotan nilai pada lirik tentunya sangat mempengaruhi hasil dari pengklasifikasian.
5. Teknik pre-processing dengan menggunakan stopword beberapa “kata” yang memiliki pembobotan data yang besar sehingga data tersebut bisa dibuang dan tidak mengganggu proses klasifikasi. Beberapa “kata” yang dimaksud yaitu 2-3 “kata” yang memiliki pembobotan paling tinggi.



Daftar Pustaka

- Afrianto, R. B., & Kurniawati, L. Y. (2013). Kategorisasi Dokuken Teks Secara Multi Label Menggunakan Fuzzy C-Means dan K-Nearest Neighbors Pada Artikel Berbahasa Indonesia. *Jurusan Teknik Informatika*, 21-26.
- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine. *Yayasan Cahaya Islam, Jurnal Teknologi Indonesia*, 1-5.
- Armianti, D. N., Indriati, & Adinugroho, S. (2019). Klasifikasi Emosi Lagu Berdasarkan Lirik pada Teks Berbahasa Indonesia. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 10161-10167.
- Astuti, L. W., C, A. R., & Lukito, Y. (2017). IMPLEMENTASI ALGORITMA NAÏVE BAYES MENGGUNAKAN ISEAR. *Jurnal Informatika*, 16-21.
- Chamunyonga, C., Edwards, C., Caldwell, P., & Burbery, J. (2020). The Impact of Artificial Intelligence and Machine Learning in Radiation. *Journal of Medical Imaging and Radiation Sciences*, 214-220.
- Chiang, T. H., LO, H. Y., & Lin, S. D. (2012). A Ranking-based KNN Approach for Multi-Label. *JMLR: Workshop and Conference Proceedings* , 81-96.
- Dillak, R. Y., Pengestuty, D. M., & Bintiri, M. G. (2012). KLASIFIKASI JENIS MUSIK BERDASARKAN FILE AUDIO. *Seminar Nasional Informatika*, c-122-c-125.
- Ghoneim, S. (2019, April 2). *towards data science*. Retrieved from towardsdatascience.com: <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>
- Gracia, S., Luengo, J., & Herrera, F. (2014). *Data Preprocessing in Data Mining*. Springer.
- Hapke, H., & Nelson, C. (2020). *Building Machine Learning Pipelines*. O'Reilly Media, Inc.

- Isnaini, N. (2019). A multi-label classification on topics of Indonesian news using K-Nearest. *Journal of Physics: Conference Series*, 1-11.
- Kusuma, D. P. (2020). *Machine Learning Teori, Program, dan Studi Kasus*. Yogyakarta: Grup Penerbitan CV BUDI UTAMA.
- Muflikhah, L., Ratnawati, D. E., & Putri, R. R. (2018). *DATA MINING*. Malang: UB Press.
- Murty, M. N., & Devi, V. S. (2015). *Introduction To Pattern Recognition And Machine Learning*. World Scientific.
- Noviardini, M. C., Osmond, A. B., & Setianingsih, C. (2018). KLASIFIKASI EMOSI PADA LIRIK LAGU MENGGUNAKAN METODE NAÏVE. *e-Proceeding of Engineering*, 6195-6202.
- Nugraha, F. A., Harani, N. H., & Habibi, R. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Bandung: Kreatif Industri Nusantara.
- Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013). Analisis Sentimen Masyarakat terhadap Calon. *JURNAL ITSMART*, 35-42.
- Prasetyowati, E. (2017). *DATA MINING Pengelompokan Data untuk Informasi dan Evaluasi*. Duta Media.
- Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 27-31.
- Rohrer, B. (2011). An Implemented Architecture for Feature. *SAND*.
- Saputri, I. S., Fadhli, M., & Surya, I. (2017). Penerapan Metode UCD (User Centered Design) pada E-Commerce Putri Intan Shop Berbasis Web. *Jurnal Nasional Teknologi dan Sistem Informasi*, 269-278.

- Saputro, P. H., Aristian, M., & Tyas, D. L. (2017). KLASIFIKASI LAGU DAERAH INDONESIA. *Jurnal Teknologi Informatika dan Terapan*, 45-50.
- Sidiq, Y. N., Fathonah, N. S., & Riza, N. (2020). *Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur Dengan Algoritma Naive Bayes Pada Sistem JURAGAN Berbasis Komputer*. Bandung: Kreatif Industri Nusantara.
- Sulartopo. (2015). PENGKATEGORIAN TOPIK SKRIPSI DENGAN METODE NBC. *JURNAL ILMIAH EKONOMI DAN BISNIS*, 49-53.
- Sun, L., Ji, S., & Ye, J. (2014). *Multi-Label Dimensionality Reduction*. Boca Raton: CRC Press Taylor & Francis Group.
- Trohidis, K., Tsoumakas, G., Kalliris, G., & Vlahavas, I. (2008). MULTI-LABEL CLASSIFICATION OF MUSIC INTO EMOTIONS. *ISMIR*, 615-620.

