

**SISTEM REKOMENDASI TOPIK SKRIPSI PRODI
INFORMATIKA**

Skripsi



**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2023

**SISTEM REKOMENDASI TOPIK SKRIPSI PRODI
INFORMATIKA**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

LUKAS KURNIAWAN

71190413

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA**

2023

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS
SECARA ONLINE
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA**

Saya yang bertanda tangan di bawah ini:

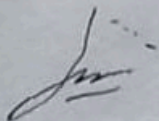
NIM : 71190413
Nama : Lukas Kurniawan
Prodi / Fakultas : Teknologi Informasi / Informatika
Judul Tugas Akhir : Sistem Rekomendasi Topik Skripsi Prodi
Informatika

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty-free Right*) serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Noneklusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk *database*, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 15 Desember 2022

Yang menyatakan,



71190413 – Lukas Kurniawan

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

SISTEM REKOMENDASI TOPIK SKRIPSI PRODI INFORMATIKA

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 kesarjanaan 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 kesarjanaan saya.

Yogyakarta, 15 Desember 2022



LUKAS KURNIAWAN
71190413

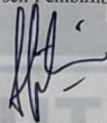
DUTA WACANA

HALAMAN PERSETUJUAN

Judul Skripsi : SISTEM REKOMENDASI TOPIK SKRIPSI PRODI
INFORMATIKA
Nama Mahasiswa : LUKAS KURNIAWAN
NIM : 71190413
Mata Kuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Ganjil
Tahun Akademik : Tahun Akademik

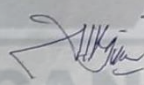
Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 15 Desember 2022

Dosen Pembimbing I



Agata Filiana, S.Kom., M.Sc.

Dosen Pembimbing II



Gloria Virginia, S.Kom., MAL., Ph.D

HALAMAN PENGESAHAN

SISTEM REKOMENDASI PEMILIHAN PROFIL LULUSAN MAHASISWA PRODI INFORMATIKA

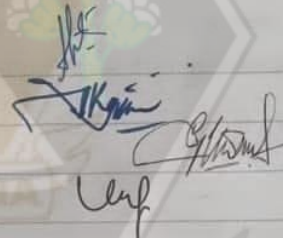
Oleh: BASTIAN SURYA HARTONO / 71190410

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 13 Januari 2023

Yogyakarta, 6 Februari 2023
Mengesahkan,

Dewan Penguji:

1. Agata Filiana, S.Kom., M.Sc.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. R. Gunawan Santosa, Drs. M.Si.
4. Matahari Bhakti Nendya, S.Kom., M.T.

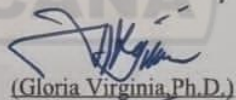


Dekan

Ketua Program Studi



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



(Gloria Virginia, Ph.D.)

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS
SECARA ONLINE
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA**

Saya yang bertanda tangan di bawah ini:


NIM : 71190413
Nama : Lukas Kurniawan
Prodi / Fakultas : Teknologi Informasi / Informatika
Judul Tugas Akhir : Sistem Rekomendasi Topik Skripsi Prodi
Informatika

bersedia menyerahkan Tugas Akhir kepada Universitas melalui Perpustakaan untuk keperluan akademis dan memberikan **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty-free Right*) serta bersedia Tugas Akhirnya dipublikasikan secara online dan dapat diakses secara lengkap (*full access*).

Dengan Hak Bebas Royalti Noneksklusif ini Perpustakaan Universitas Kristen Duta Wacana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk *database*, merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Yogyakarta, 15 Desember 2022

Yang menyatakan,


71190413 – Lukas Kurniawan



Karya sederhana ini dipersembahkan

kepada Tuhan, Keluarga Tercinta,

dan Kedua Orang Tua



Orang gagal bukan karena dia tidak berhasil, tetapi orang gagal karena dia tidak pernah mau mencoba.

Anonim

If you aren't confused by quantum mechanics, you haven't really understood it.

Niels Bohr

KATA PENGANTAR

Segala puji dan syukur kepada Tuhan yang maha kasih, karena atas segala rahmat, bimbingan, dan bantuan-Nya maka akhirnya Skripsi dengan judul Sistem Rekomendasi Topik Skripsi Prodi Informatik 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,
2. Orang tua yang selama ini telah sabar membimbing dan mendoakan penulis tanpa kenal untuk selama-lamanya,
3. Bapak Restyandito, S.Kom, MSIS., Ph.D. selaku Fakultas Teknologi Informasi UKDW.
4. Ibu Gloria Virginia, S.Kom., MAI., Ph.D. selaku Ketua Program Studi Informatika Fakultas Teknologi Informasi UKDW.
5. Agata Filiana, S.Kom., M.Sc. selaku Dosen Pembimbing 1, yang telah memberikan ilmunya dan dengan penuh kesabaran membimbing penulis.
6. Ibu Gloria Virginia, S.Kom., MAI., Ph.D. selaku Dosen Pembimbing 2 yang telah memberikan ilmu dan kesabaran dalam membimbing penulis.
7. Keluarga tercinta yang telah membantu saya dari segala aspek demi kesuksesan skripsi ini.
8. Teman-teman 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, 15 Desember 2022

Lukas Kurniawan



DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS SECARA ONLINE.....	vi
UNIVERSITAS KRISTEN DUTA WACANA YOGYAKARTA	vi
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
INTISARI.....	xix
ABSTRACT.....	xxi
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Sistematika Penulisan.....	3
BAB II.....	5
TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori	9
2.2.1 Sistem Rekomendasi	9
2.2.2 K-Means	10
2.2.3 K-Means++	12
2.2.4 Principal component analysis (PCA)	13

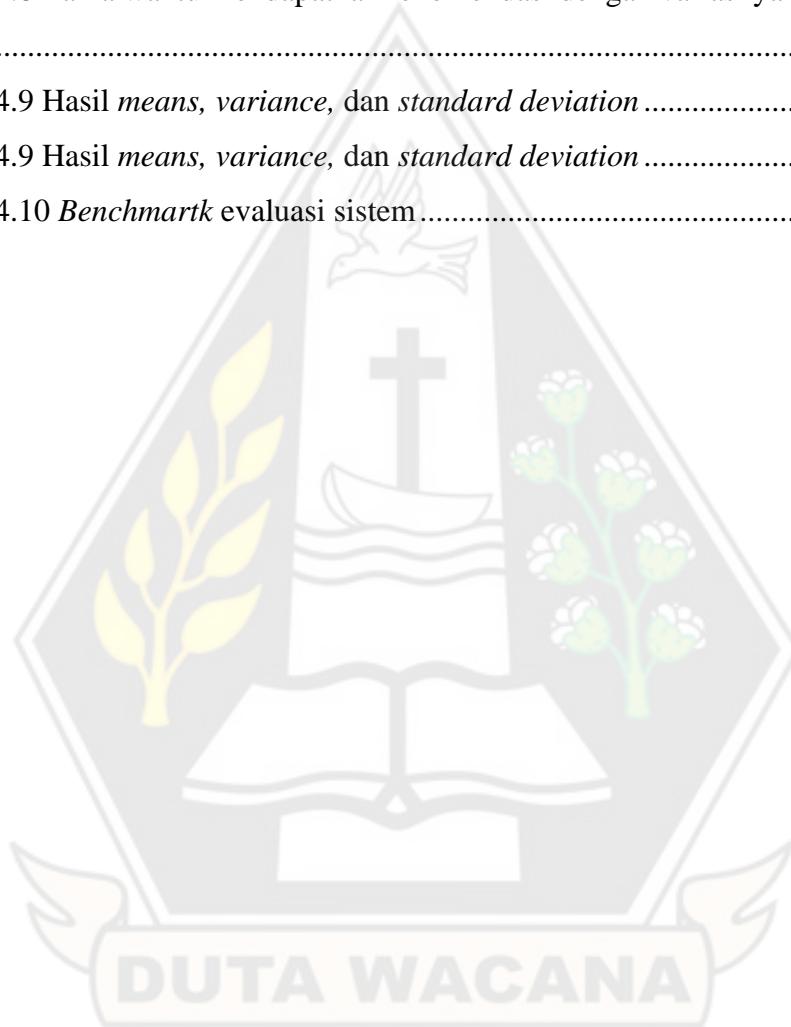
2.2.5	Latent Dirichlet Allocation (LDA).....	14
2.2.6	<i>Term Frequency Inverse Document Frequency</i>	16
2.2.7	Cosine similarity	17
2.2.8	Evaluasi	18
2.2.9	Pengujian.....	20
2.2.10	UEQ (<i>User Experience Questionnaire</i>)	21
2.2.11	Rumus Slovin	22
BAB III		24
METODOLOGI PENELITIAN.....		24
3.1	Analisis Kebutuhan Sistem	24
3.2	Perancangan Penelitian.....	26
3.3	Alur Pengembangan Model.....	28
3.4	Alur Sistem Rekomendasi Topik Skripsi	31
3.5	<i>Use Case</i> Diagram Sistem	32
3.6	Arsitektur Sistem	34
3.7	Perancangan Basis Data	35
3.8	Perancangan Evaluasi dan Pengujian Model.....	36
3.9	Pengujian Sistem	37
3.10	Contoh Kasus	39
BAB IV		72
IMPLEMENTASI DAN PEMBAHASAN.....		72
4.1	Implementasi Sistem	72
4.1.1	Antarmuka Sistem.....	72
4.1.2	Implementasi pengumpulan data.....	78
4.1.3	Implementasi persiapan data nilai	81
4.1.4	Implementasi persiapan data skripsi	94
4.1.5	Implementasi Model K-Means++	106
4.1.6	Implementasi <i>Cosine Similarity</i>	110

4.1.7	Implementasi LDA Gibbs Sampling.....	114
4.1.8	Implementasi penyamaan kata untuk <i>query</i>	118
4.2	Hasil, Pengujian dan Analisa Sistem.....	118
4.2.1.	Hasil Sistem dan fitur filterisasi.....	118
4.2.2.	Pengujian dan Analisa Sistem.....	120
BAB V	130
KESIMPULAN DAN SARAN	130
5.1	Kesimpulan.....	130
5.2	Saran.....	131
DAFTAR PUSTAKA	132
LAMPIRAN A	136
KODE SUMBER PROGRAM	136
LAMPIRAN B	202
KARTU KONSULTASI DOSEN 1	202
LAMPIRAN C	203
KARTU KONSULTASI DOSEN 2	203
LAMPIRAN D	204
LAMPIRAN LAIN-LAIN	204
D.1.	Tabel data kuesioner 2019 dan 2020.....	204
D.2.	Tabel data kuesioner 2019 dan 2020 yang ditransformasi.....	205
D.3.	Tabel data kuesioner 2019.....	206
D.4.	Tabel data kuesioner 2019 yang ditransformasi.....	207
D.5.	Tabel data kuesioner 2020.....	207
D.6.	Tabel data kuesioner 2020 yang ditransformasi.....	208

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	20
Tabel 2.2 <i>Benchmark</i> Interval skala UEQ.....	22
Tabel 3.1 Daftar Kebutuhan Pengguna	25
Tabel 3.2 Daftar Kebutuhan Pengguna	26
Tabel 3.3 <i>Use Case Log in</i>	33
Tabel 3.4 <i>Use case</i> Lihat rekomendasi pada 1 mahasiswa	33
Tabel 3.5 <i>Use case</i> Lihat rekomendasi pada 1 mahasiswa	34
Tabel 3.6 Data contoh untuk K-Means++.....	40
Tabel 3.7 Nilai rata-rata setiap feature	41
Tabel 3.8 Data yang telah dikurangi dengan nilai rata-rata nya	41
Tabel 3.9 Nilai <i>covariance matrix</i> pada setiap <i>feature</i>	42
Tabel 3.10 Hasil akhir proses PCA dari data nilai mahasiswa perhitungan manual dibandingkan dengan Scikit-learn.....	45
Tabel 3.11 Hasil akhir klasterisasi K-Means++ untuk K=1 dengan perhitungan manual dan Scikit-learn.....	48
Tabel 3.12 Hasil akhir klasterisasi K-Means++ untuk K=2.....	50
Tabel 3.13 Pembobotan dokumen dengan menggunakan persamaan[3] dan persamaan[4].....	57
Tabel 3.14 Pembobotan <i>query</i> menggunakan persamaan[3] dan persamaan[4]... ..	58
Tabel 3.15 Hasil pembobotan dokumen dan <i>query</i> menggunakan Library scikit-learn pada Python.....	58
Tabel 3.16 <i>Cosine similarity</i> dokumen 1 terhadap <i>query</i>	59
Tabel 3.17 <i>Cosine similarity</i> dokumen 2 terhadap <i>query</i>	59
Tabel 4.1 Perbandingan nilai <i>cosine similarity</i> sebelum dengan sesuah pra-pemrosesan 2.....	111
Tabel 4.2 Hasil presisi dan <i>recall</i>	112
Tabel 4.3 Hasil <i>perplexity</i> topik 1,5,10,15,20,25,30, dan 35	115
Tabel 4.4 Hasil <i>perplexity</i> topik 31,32,33,34, dan 35	115
Tabel 4.5 Daftar lama waktu permodelan topik, untuk iterasi 10 dan 20	116

Tabel 4.5 Daftar lama waktu permodelan topik, untuk iterasi 10 dan 20	117
Tabel 4.6 Lama permodelan topik pada data skripsi setiap kluster.....	117
Tabel 4.7 Hasil penyamaan <i>query</i> ke dalam <i>term</i> yang sama pada dokumen.....	118
Tabel 4.8 Lama waktu mendapatkan rekomendasi dengan variasi yang diberikan	120
Tabel 4.8 Lama waktu mendapatkan rekomendasi dengan variasi yang diberikan	121
Tabel 4.9 Hasil <i>means</i> , <i>variance</i> , dan <i>standard deviation</i>	122
Tabel 4.9 Hasil <i>means</i> , <i>variance</i> , dan <i>standard deviation</i>	123
Tabel 4.10 <i>Benchmark</i> evaluasi sistem.....	124



DAFTAR GAMBAR

Gambar 2.1 Klasterisasi buruk K-Means	12
Gambar 2.2 Klasterisasi baik K-Means	12
Gambar 2.3 Permodelan Topik	14
Gambar 3.1 Metodologi Penelitian	26
Gambar 3.2 Gambaran Cara Rekomendasi Topik	27
Gambar 3.3 <i>Flowchart</i> pengembangan model	28
Gambar 3.4 Mempersiapkan data skripsi	30
Gambar 3.5 Mempersiapkan data nilai	31
Gambar 3.6 <i>Flowchart</i> pengguna sistem	32
Gambar 3.7 <i>Use Case Diagram</i> Sistem	33
Gambar 3.8 Arsitektur sistem rekomendasi topik skripsi	34
Gambar 3.9 Skema basis data sistem	35
Gambar 3.10 Daftar pertanyaan UEQ	38
Gambar 3.11 Visualisasi data Tabel 3.4 sebelum di ditengahkan	41
Gambar 3.12 Visualisasi data yang sudah di tengahkan	41
Gambar 3.13 Visualisasi <i>eigenvector</i> pada data latih yang sudah di tengahkan... ..	44
Gambar 3.14 PCs Ke-1	44
Gambar 3.15 PCs ke-2	44
Gambar 3.16 Visualisasi K-Means++ dengan K=1	48
Gambar 3.17 Visualisasi K-Means++ dengan K=2	51
Gambar 3.18 Visualisasi K-Means++ untuk K=1 menggunakan Library scikit-learn	51
Gambar 3.19 Visualisasi K-Means++ untuk K=2 menggunakan Library scikit-learn	52
Gambar 3.20 SSE untuk jumlah klaster 1-4 menggunakan Library scikit-learn pada Python	54
Gambar 4.1 Halaman <i>log in</i>	72
Gambar 4.2 Halaman <i>home</i>	72
Gambar 4.3 Halaman rekomendasi topik skripsi dengan informasi	73

Gambar 4.4 Halaman rekomendasi topik skripsi tanpa informasi	74
Gambar 4.5 Halaman proses	74
Gambar 4.6 Halaman hasil rekomendasi topik	74
Gambar 4.7 Topik dari kata “data” adalah 9.....	75
Gambar 4.8 <i>Bar chart</i> komparasi nilai mahasiswa pencari dengan rata-rata nilai mahasiswa yang skripsinya dijadikan acuan rekomendasi topik	75
Gambar 4.9 Daftar mata kuliah yang sudah diambil oleh mahasiswa pencari, dibandingkan dengan mahasiswa yang dijadikan acuan rekomendasi topik skripsi.	76
Gambar 4.10 Daftar skripsi yang dijadikan acuan memberikan topik skripsi	76
Gambar 4.11 Daftar topik skripsi yang direkomendasikan.....	76
Gambar 4.12 Detail informasi terhadap setiap topik skripsi.....	77
Gambar 4.13 Tombol <i>Dropdown</i>	78
Gambar 4.14 <i>Log out</i> sistem	78
Gambar 4.15 Urutan <i>web scraping</i>	79
Gambar 4.16 Halaman laman yang dijadikan <i>root</i>	79
Gambar 4.17 Daftar Skripsi pada tahun 2022.....	80
Gambar 4.18 Salah satu isi data skripsi	80
Gambar 4.19 Proses penghilangan redundansi 1	83
Gambar 4.20 Proses Penghilangan redundansi 2	84
Gambar 4.21 Pencatatan manual terhadap mata kuliah redundansi 3.....	90
Gambar 4.22 Visualisasi 20 jumlah <i>term</i> terbanyak pada seluruh dokumen.....	104
Gambar 4.23 Visualisasi 20-100 jumlah <i>term</i> terbanyak pada seluruh dokumen.....	104
Gambar 4.24 Visualisasi 100 jumlah <i>term</i> terbanyak dari seluruh dokumen	105
Gambar 4.25 Bentuk data nilai yang dijadikan sebagai data pelatihan model....	106
Gambar 4.26 Nilai SSE pada setiap percobaan jumlah kluster	107
Gambar 4.27 Nilai K=1	108
Gambar 4.28 Nilai K=2.....	108
Gambar 4.29 Nilai K=3.....	108
Gambar 4.30 Nilai K=4.....	108
Gambar 4.31 Nilai K=5.....	108

Gambar 4.32 Nilai K=6.....	108
Gambar 4.33 Visualisasi nilai SSE pada model yang tidak memakai PCA.....	109
Gambar 4.34 Nilai K=1	109
Gambar 4.35 Nilai K=2.....	109
Gambar 4.36 Nilai K=3.....	110
Gambar 4.37 Nilai K=4.....	110
Gambar 4.38 Nilai K=5.....	110
Gambar 4.39 Nilai K=6.....	110
Gambar 4.40 Nilai <i>perplexity</i> untuk topik 1-10	114
Gambar 4.41 Nilai <i>perplexity</i> untuk topik 11-20	114
Gambar 4.42 Nilai <i>perplexity</i> data skripsi pada klaster 1	115
Gambar 4.43 Nilai <i>perplexity</i> data skripsi pada klaster 2	115
Gambar 4.44 Nilai <i>perplexity</i> data skripsi pada klaster 3	116
Gambar 4.45 Nilai <i>perplexity</i> data skripsi pada klaster 4	116
Gambar 4.46 Persebaran data responden berdasarkan tahun angkatan.....	122
Gambar 4.47 Visualisasi nilai enam skala UEQ	123
Gambar 4.48 <i>Benchmark</i> evaluasi sistem	124
Gambar 4.49 rata-rata nilai 6 skala UEQ mahasiswa angkatan 2019	125
Gambar 4.50 Nilai rata-rata 6 skala UEQ pada tahun angkatan 2020.	125
Gambar 4.51 <i>Benchmark</i> evaluasi sistem untuk mahasiswa angkatan 2019	126
Gambar 4.52 <i>Benchmark</i> evaluasi sistem untuk mahasiswa angkatan 2020	126
Gambar 4.53 Apakah mahasiswa sudah memiliki gambaran topik skripsi yang ingin diambil?	127
Gambar 4.54 Apakah mahasiswa mendapatkan rekomendasi topik dengan memakai filterisasi “nilai miri panda”?.....	127
Gambar 4.55 Apakah visualisasi memudahkan dalam memahami rekomendasi topik yang diberikan?.....	128

INTISARI

SISTEM REKOMENDASI TOPIK SKRIPSI PRODI INFORMATIKA

Oleh

LUKAS KURNIAWAN

71190413

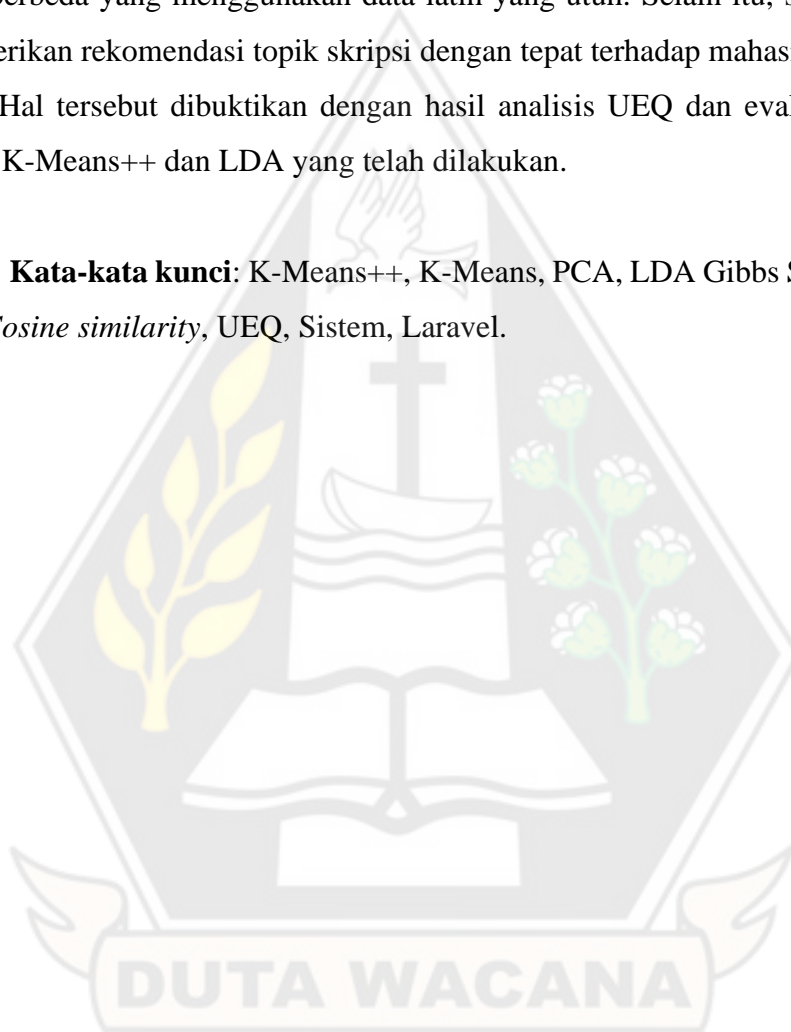
Salah satu syarat kelulusan kuliah adalah lulus skripsi. Pada skripsi, mahasiswa perlu menentukan topik skripsi. Topik skripsi yang ditentukan, menjadi dasar untuk membuat judul skripsi. Penentuan topik adalah hal yang penting pada skripsi, karena topik yang tepat akan mengurangi kendala dalam membuat skripsi. Penulis melakukan analisa terhadap persentase nilai E pada nilai mata kuliah seminar dan skripsi. Hasil analisis menunjukkan terdapat 21.6% dari 351 mahasiswa seminar yang mendapat nilai E dan 61.25% dari 240 mahasiswa yang belum dapat lulus skripsi. Hal tersebut menunjukkan mahasiswa belum memiliki kesiapan dalam mengerjakan skripsi.

FTI UKDW (Fakultas Teknologi Informasi Universitas Kristen Duta Wacana) ingin membuat sistem rekomendasi topik skripsi agar mahasiswa dapat memilih topik skripsi dengan tepat. Sistem rekomendasi topik skripsi dibangun menggunakan *data warehouse* FTI UKDW. Sistem rekomendasi topik skripsi dibuat menggunakan kombinasi antara *K-Means++*, *PCA (principal component analysis)*, *cosine similarity* dan *LDA (Latent Dirichlet Allocation) Gibbs sampling*. Sistem rekomendasi diimplementasikan ke dalam suatu *website* menggunakan *Laravel*. *Website* dapat diakses secara *online* oleh mahasiswa menggunakan perangkat komputer atau laptop.

Hasil penelitian, menunjukkan sistem rekomendasi topik skripsi dapat dibuat menggunakan kombinasi *K-Means++*, *PCA*, *cosine similarity*, dan *LDA*

Gibbs sampling. Model K-Means++ menghasilkan 4 kluster yang didapat dari analisa menggunakan *elbow method*. LDA Gibbs sampling diimplementasikan dengan iterasi maksimal sebanyak 10 dan jumlah topik sebanyak 20 serta jumlah maksimal dokumen yang dapat diproses adalah 29. Penggunaan PCA untuk mengurangi jumlah dimensi pada data dapat menghasilkan model K-Means++ yang tidak berbeda yang menggunakan data latih yang utuh. Selain itu, sistem mampu memberikan rekomendasi topik skripsi dengan tepat terhadap mahasiswa angkatan 2019. Hal tersebut dibuktikan dengan hasil analisis UEQ dan evaluasi terhadap model K-Means++ dan LDA yang telah dilakukan.

Kata-kata kunci: K-Means++, K-Means, PCA, LDA Gibbs Sampling, TF-IDF, *Cosine similarity*, UEQ, Sistem, Laravel.



ABSTRACT

RECOMMENDATION SYSTEM TOPIC THESIS STUDY PROGRAM INFORMATICS

By

LUKAS KURNIAWAN

71190413

One of the requirements for graduating from college is passing a thesis. In thesis, students need to determine the topic of thesis. The specified thesis topic becomes the basis for making the thesis title. Determining a topic is important in a thesis, because the right topic will reduce obstacles in writing a thesis (Gema & Ahyanuardi, 2021). The author analyzes the percentage of E scores in seminar and thesis grades. The results of the analysis showed that there were 21.6% of the 351 seminar students who received an E grade and 61.25% of the 240 students who could not pass their thesis. This shows that students do not have the readiness to work on thesis.

UKDW FTI (Faculty of Information Technology, Duta Wacana Christian University) wants to create a thesis topic recommendation system so that students can choose the right thesis topic. The thesis topic recommendation system was built using the UKDW FTI data warehouse. Thesis topic recommendation system is made using a combination of K-Means++, cosine similarity and LDA (Latent Dirichlet Allocation) Gibbs sampling. The recommendation system is implemented into a website using Laravel. The website can be accessed online by students using a computer or laptop device.

The results showed that the thesis topic recommendation system could be made using a combination of K-Means++, PCA, cosine similarity, and LDA Gibbs sampling. The K-Means++ model produces 4 clusters which are obtained from the analysis using the elbow method. LDA Gibbs sampling is implemented with a

maximum iteration of 10 and the number of topics is 20 and the maximum number of documents that can be processed is 29. The use of PCA to reduce the number of dimensions in the data can produce K-Means++ models that are not different using the intact training data. In addition, the system is able to provide appropriate thesis topic recommendations for class 2019 students. This is evidenced by the results of the UEQ analysis and evaluation of the K-Means++ and LDA models that have been carried out.

Keywords: K-Means++, K-Means, PCA, LDA Gibbs Sampling, TF-IDF, *Cosine similarity*, UEQ, System, Laravel.



BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Salah satu syarat kelulusan kuliah adalah lulus skripsi. Pada skripsi, mahasiswa perlu menentukan topik skripsi. Topik skripsi yang ditentukan, menjadi dasar untuk membuat judul skripsi. Penentuan topik adalah hal yang penting pada skripsi, karena topik yang tepat akan mengurangi kendala dalam membuat skripsi (Gema & Ahyanuardi, 2021).

Proses analisa pada *data warehouse* FTI menunjukkan terdapat mahasiswa yang mengulang mata kuliah Riset Teknologi Informasi, Seminar, dan Skripsi. Mata kuliah RTI (Riset Teknologi Informasi) dan Seminar adalah dua mata kuliah yang disediakan untuk mempersiapkan Skripsi. Hasil analisis menunjukkan 3% dari 378 mahasiswa yang mendapat nilai E pada RTI. Pada mata kuliah seminar, terdapat 21.5% dari 351 mahasiswa yang mendapatkan nilai E. Sedangkan, pada mata kuliah skripsi terdapat 39.1% dari 240 mahasiswa yang belum lulus. Semua hasil analisis tersebut dilakukan pada data nilai dengan kondisi mahasiswa telah mengulang pada mata kuliah yang diulang. Hasil analisis terhadap data nilai yang belum terdapat pengulangan menunjukkan persentase yang lebih tinggi, yaitu 4% untuk RTI, 21.6% untuk Seminar, dan 61.25% untuk Skripsi. Pada setiap mata kuliah di atas, rata-rata mahasiswa mengulang sebanyak 2-3 kali. Hasil analisis tersebut menunjukkan cukup banyak mahasiswa yang belum mempersiapkan skripsi dengan baik.

FTI UKDW (Fakultas Teknologi Informasi Universitas Kristen Duta Wacana) ingin membuat sistem rekomendasi topik skripsi agar mahasiswa dapat memilih topik skripsi dengan tepat. Sistem rekomendasi topik skripsi dibangun menggunakan *data warehouse* FTI UKDW. Sistem rekomendasi topik skripsi dibuat menggunakan kombinasi antara *K-Means++*, *cosine similarity* dan LDA

(*Latent Dirichlet Allocation*) Gibbs sampling. Mahasiswa yang lulus skripsi dapat diklasterisasi memakai K-Means++ berdasarkan data mata kuliah dan nilainya. Proses tersebut dapat dipakai untuk menentukan klaster dari mahasiswa pencari topik skripsi. Data skripsi yang berasal dari klaster mahasiswa pencari digunakan untuk proses selanjutnya. Kemudian, mahasiswa pencari memasukan *input* untuk mencari data skripsi yang relevan. Proses tersebut dilakukan menggunakan *cosine similarity*. Data skripsi yang didapat dilakukan permodelan topik menggunakan LDA Gibbs sampling sehingga menghasilkan topik skripsi. Sistem rekomendasi diimplementasikan ke dalam suatu *website* menggunakan Laravel. *Website* dapat diakses secara *online* oleh mahasiswa menggunakan perangkat komputer atau laptop. Penelitian ini memiliki harapan agar setiap mahasiswa memperoleh topik skripsi yang sesuai dengan keinginannya, sehingga mahasiswa dapat dengan mudah membuat judul skripsi yang sesuai dengan topik yang dikuasai.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang ada, maka disusun rumusan masalah sebagai berikut:

1. Bagaimana cara membangun sistem rekomendasi topik skripsi dengan menggunakan kombinasi K-Means++, *cosine similarity*, LDA?
2. Apakah sistem rekomendasi topik skripsi dapat memberikan rekomendasi yang tepat bagi mahasiswa?

1.3. Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah seperti berikut:

1. Data mahasiswa Prodi Informatika yang dijadikan sebagai bahan pelatihan dan pengujian didapat dari tahun 2015 hingga tahun 2021.
2. Dari tahun 2015 hingga 2021 telah terjadi perubahan kurikulum sebanyak 4 kali. Perubahan kurikulum memberikan dampak dengan perubahan beberapa aturan perkuliahan. Beberapa aturan di kurikulum 2011 Revisi

kurikulum 2013, kurikulum 2015, kurikulum 2017 dan kurikulum 2017 revisi 2019 akan dikonversi menjadi aturan kurikulum 2021 dengan penyesuaian.

3. Data nilai yang dipakai diperbarui pada tahun pembelajaran 2021/2022 semester gasal.
4. Pengguna sistem adalah mahasiswa semester 6 karena memiliki jumlah data mata kuliah yang mencukupi untuk memprediksi topik skripsi.

1.4. Tujuan Penelitian

Penelitian bertujuan untuk membuat sistem rekomendasi topik skripsi yang mampu memberikan rekomendasi tepat dengan menggunakan kombinasi K-Means++, LDA Gibbs sampling, dan *cosine similarity*.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini dirasakan oleh 3 pihak, yaitu Fakultas Teknologi Informasi UKDW, Dosen wali, dan Mahasiswa. Mahasiswa yang akan mengambil atau mengulang skripsi dapat dipermudah memilih topik skripsi. Karena mahasiswa dapat mengakses *website* yang memberikan rekomendasi topik skripsi sesuai dengan mata kuliah yang diambil dan nilainya. Dosen wali dapat memberikan arahan mengenai judul skripsi, terhadap mahasiswa yang melakukan konsultasi dari informasi rekomendasi topik skripsi yang didapat. Fakultas Informatika UKDW memiliki keuntungan jika setiap mahasiswa dapat lulus skripsi tepat waktu. Karena akan memperbesar peluang mahasiswa untuk lulus dari perkuliahan tepat waktu.

1.6. Sistematika Penulisan

Pada tugas akhir ini terdapat lima bab yang disajikan antara lain:

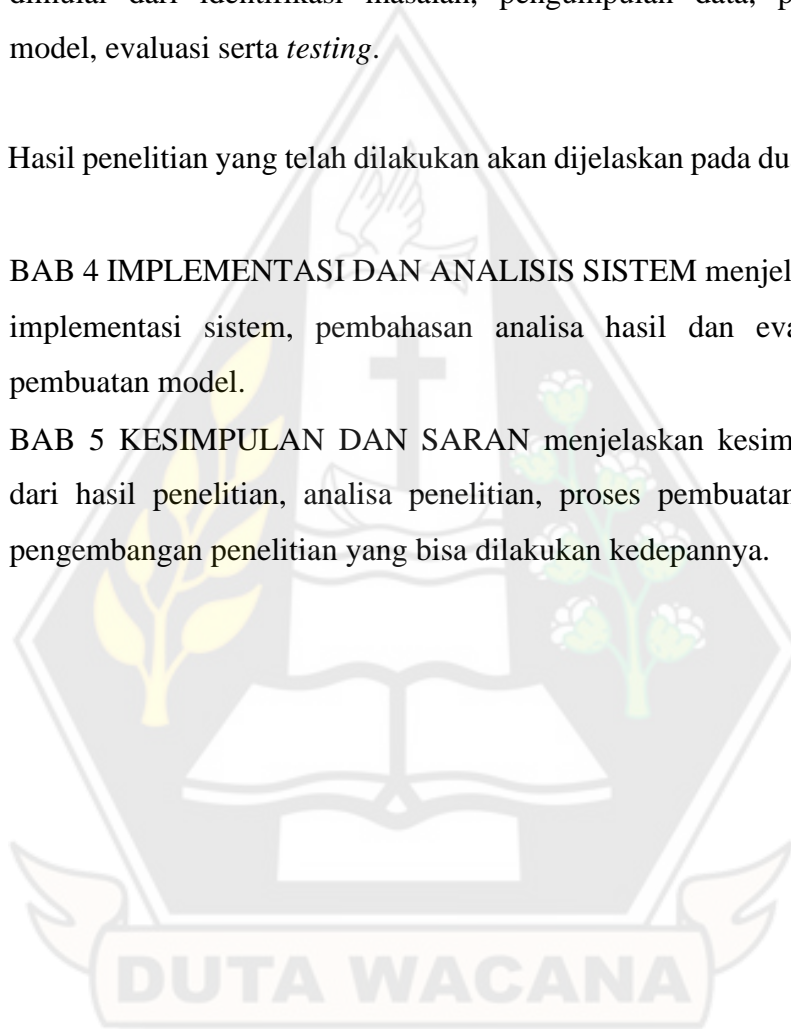
- BAB 1 PENDAHULUAN berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan laporan.
- BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI berisi mengenai kumpulan teori dan tinjauan yang digunakan dalam mendukung

penelitian. Teori atau tinjauan yang ada terdiri dari definisi, konsep, permodelan, persamaan atau pendapat para ahli, dan hasil penelitian yang didapat dari pustaka.

- **BAB 3 ANALISIS DAN PERANCANGAN SISTEM** berisi mengenai tahapan yang dilakukan dalam membangun sistem rekomendasi yang dimulai dari identifikasi masalah, pengumpulan data, pengembangan model, evaluasi serta *testing*.

Hasil penelitian yang telah dilakukan akan dijelaskan pada dua bab terakhir, yaitu:

- **BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM** menjelaskan tahapan implementasi sistem, pembahasan analisa hasil dan evaluasi selama pembuatan model.
- **BAB 5 KESIMPULAN DAN SARAN** menjelaskan kesimpulan singkat dari hasil penelitian, analisa penelitian, proses pembuatan model serta pengembangan penelitian yang bisa dilakukan kedepannya.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Sistem rekomendasi topik skripsi berhasil dibangun dan memberikan rekomendasi yang tepat menggunakan kombinasi model yang telah dirancang. Selain itu, penelitian ini menghasilkan kesimpulan sebagai berikut:

1. Sistem rekomendasi topik skripsi memakai model K-Means++ dengan jumlah kluster 4. Jumlah tersebut didapati berdasarkan proses evaluasi menggunakan *elbow method* dan visualisasi klusterisasi data. Klusterisasi Model K-Means++ akan mengelompokan data skripsi berdasarkan nilai mata kuliah pembuatnya. *Cosine similarity* digunakan untuk mencari dokumen skripsi dari kluster hasil prediksi yang relevan dengan *query* pencarian. LDA Gibbs sampling digunakan untuk membuat topik dengan jumlah dari 1 hingga 20. Jumlah dokumen maksimal yang dijadikan permodelan topik adalah 29. Setiap pembuatan topik akan melakukan iterasi maksimal sebanyak 10 dan pada iterasi terakhir akan dicatat nilai *perplexity*. Jumlah topik dengan nilai *perplexity* terendah menandakan jumlah topik yang ideal untuk dihasilkan.
2. Analisis *user experience* menggunakan UEQ menunjukkan bahwa sistem dapat memberikan rekomendasi topik skripsi yang tepat terhadap mahasiswa angkatan 2019. Hal tersebut disebabkan karena nilai 6 skala UEQ yang didapat dari angkatan 2019 memiliki nilai yang baik. Sedangkan pada 6 mahasiswa angkatan 2020 sistem kurang dapat memberikan rekomendasi dengan tepat. Hal ini ditunjukkan dengan nilai skala ketepatan yang bernilai jelek.
3. Pyscript untuk menjalankan *script* Python memberikan dampak buruk dengan lamanya proses eksekusi *script* Python di awal proses.
4. Proses filterisasi dokumen skripsi dengan menggunakan klusterisasi K-Means++ berhasil memberikan rekomendasi topik skripsi yang sesuai

dengan mahasiswa. Hal ini dibuktikan pada Gambar 4.54, dimana mayoritas mendapatkan rekomendasi topik skripsi dengan menggunakan filterisasi nilai yang mirip.

5.2 Saran

Penulis memiliki beberapa saran terhadap studi penelitian dan pengembangan terhadap sistem yang dapat dilakukan pada kemudian hari:

1. *Framework* yang digunakan dalam membangun sistem dapat menggunakan Django yang dirancang untuk dapat mengeksekusi *script* Python.
2. Pencatatan *term* yang memiliki kemungkinan variasi penulisan dan keterangan perlu dilakukan dengan lebih detail dan rapi.
3. Penentuan nilai parameter *alpha* dan *beta* perlu diteliti pada penelitian selanjutnya.
4. Kuesioner perlu dilakukan pada jangka waktu yang lebih Panjang, karena jumlah responden tidak mencapai target yang ditentukan. Selain itu perlu adanya orang dengan pengaruh besar untuk bisa mengajak orang dalam pengumpulan responden dari berbagai tahun angkatan.
5. Implementasi LDA ke Python dapat menggunakan library lain yang sudah diketahui, seperti scikit-learn dan Gensim.
6. Sistem dapat dilengkapi dengan tahapan yang urut yang menjelaskan cara sistem bekerja. Selain itu, informasi yang berupa visualisasi dapat dilengkapi dengan penjelasan agar pengguna dapat memahami dengan mudah cara menggunakan sistem dan memahami informasi yang diberikan.

DAFTAR PUSTAKA

- Afifuddin, R. N., & Nurjanah, D. (2019). Sistem Rekomendasi Pemilihan Mata kuliah Peminatan Menggunakan Algoritma K-means dan Apriori (studi kasus: Jurusan S1 Teknik Informatika Fakultas Informatika)., (pp. 2359-2367).
- Alfanzar, A. I., Khalid, & Rozas, I. S. (2020, March). Topic Modelling Skripsi Menggunakan Metode Latent Dirichlet Allocation. *Jurnal Sistem Informasi*, 7(1), 7-13.
- Anbazhagan, N. (2011). Improving the Performance of K-Means. *International Journal on Computer Science and Engineering (IJCSE)*, 3(6), 2317-2322.
- Apriani, Zakiyudin, H., & Marzuki, K. (2021, June 1). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, 3(1), 19-27.
- Aubaidan, B., Mohd, M., & Albared, M. (2014). Comparative Study of K-means and K-means++ Clustering on Crime Dromain. *Journal of Computer Science*, 10(7), 1197-1206.
- Barber, S. (2009). *How Fast Does a Website Need To Be?* . Retrieved from http://www.perftestplus.com/resources/how_fast.pdf
- Bashiri, M. F. (2017). Analisa Sentimen Menggunakan Latent Dirichlet Allocation dan Visualisasi Topic Popularity Wordcloud. Ungaran. Retrieved from <https://core.ac.uk/download/pdf/151236958.pdf>
- Boyd-Graber, J. (2017). *Applications of Topic Models*. Boulder: now publishers.
- Francis, B. K., & Babu, S. S. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems*, 43, 162.
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi dan Sistem Informasi* , 82-89.

- Haryoso, A. (2017). Analisis Website Performance Milik Kementerian di Indonesia Menggunakan Metode Pembobotan Entropi Dan Metode Pemeringkatan Electre. Yogyakarta: <https://eprints.uny.ac.id/54806/>. Retrieved from <https://eprints.uny.ac.id/54806/1/Skripsi.pdf>
- Haviluddin, Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means untuk Pengelompokan Rekomendasi. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*.
- Henim, S. R., & Sari, R. P. (2020). Evaluasi User Experience Sistem Informasi Akademik Mahasiswa pada Perguruan Tinggi Menggunakan User Experience Questionnaire. *Jurnal Komputer Terapan*, 69–78.
- Isinkaye, F. O., Yolajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 261-273.
- Janura, G., & Ahyanuardi. (2021). Analisis Kendala Mahasiswa dalam Penyelesaian Skripsi pada Masa Pandemi Covid-19. *Jurnal Pendidikan Teknik Elektro*.
- Jepriana, I. W., & Hanief, S. (2020). Metode item-based Collaborative Filtering Untuk Model Sistem Rekomendasi Konsentrasi Penjurusan di STMIK STIKOM Bali. *Jurnal Teknologi Informasi dan Komputer*.
- Kurniadi, D., Haviana, S. F., & Novianto, A. (2020). Implementasi Algoritma Cosine Similarity pada sistem arsip dokumen di Universitas Islam Sultan Agung. *TRANSFORMTIKA*, 124-132.
- Merliana, N. P., Ernawati, & Santoso, A. J. (n.d.). Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU*. Yogyakarta.
- Nugraha, A. Y., & Abdulloh, F. F. (2022). Optimasi Naive Bayes dan Cosine Similarity Menggunakan Particle. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 1444-1451.

- Nugroho, N., & Adhinata, F. D. (2022). Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa. *TEKNIKA*, 11(3), 170-179.
- Nurlayli, A., & Nasichuddin, M. A. (2019). Topic Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation. *ELINVO (Electronics, Informatics, and Vocational Education)*, 4(2), 154-161.
- Prabhu, P., & Anbazhagan, N. (2011). Improving the Performance of K-Means Clustering For High Dimensional Data Set. *International Journal on Computer Science and Engineering (IJCSE)*, 3(6), 2317-2322.
- Purwitasari, D., Muflichah, A., Hasanah, N. A., & Arifin, A. Z. (2021). Pemodelan Topik dengan LDA untuk Temu Kembali Informasi dalam Rekomendasi Tugas Akhir. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 5(3), 421-428.
- Putra, I. M., & Kusumawardani, n. R. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *JURNAL TEKNIK ITS*, 6(2).
- Rahman, A., Waskitho, R. B., Nuha, M. F., & Rakhmawati, N. A. (2021). Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation. *Journal Information Engineering and Educational Technology*, 5(2).
- Riyani, A., Nafan, M. Z., & Hanuddin, A. B. (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen. *Journal Linguistik Komputasional*, 2(1).
- Rukmi, A. M., & Iqbal, I. M. (2017). Using k-means++ algorithm for researches clustering. AIP Publishing.
- Sahria, Y., & Fudholi, D. H. (2020). Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation). *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 4(2), 336–344.

- Santika, T. (2019). *Evaluasi Perplexity Untuk Pemodelan Topik Diskusi Agama Islam di Media Sosial Twitter Indonesia Tahun 2006-2018 Menggunakan Latent Dirichlet Allocation*. Jakarta.
- Sari, R. Y., Oktavianto, H., & Sulistyono, H. W. (2022). Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia. *Jurnal Smart Teknologi*, 3(2), 104-108.
- Toraismaya, A., Sasongko, L. R., & Rondonuwu, F. S. (2020). Principal Component Dan K-Means Cluster Analysis Untuk Data Spektrum Black Tea Grades Guna Penilaian Kualitas Alternatif. *JOURNAL OF FUNDAMENTAL MATHEMATICS AND APPLICATIONS (JFMA)*, 3(2).
- Wijaya, I. N., Iswara, I. B., & Arsana, I. N. (2021). Analisis dan Evaluasi Pengalaman Pengguna PaTik Bali Dengan Metode User Experience Questionnaire (UEQ). *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 8(2), 217-226.

