

**CLUSTERING JAWABAN URAIAN MAHASISWA
MENGUNAKAN TF-IDF DAN KMEANS**

SKRIPSI

Digunakan Sebagai Syarat Maju Ujian Diploma IV
Politeknik Negeri Malang

Oleh:

IRFAN THALIB ALFARID

NIM. 1941720039



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
JULI 2023**

HALAMAN PENGESAHAN

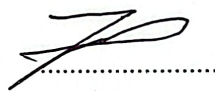



CLUSTERING JAWABAN URAIAN MAHASISWA MENGUNAKAN TF-IDF DAN KMEANS

Disusun Oleh:

IRFAN THALIB ALFARID NIM. 1941720039

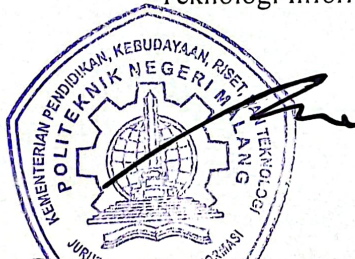
Laporan Akhir ini telah diuji pada tanggal 26 Juli 2023

Disetujui oleh:

- | | | |
|-----------------------------|--|---|
| 1. Pembimbing
Utama | : <u>Irsyad Arif Mashudi, S.Kom., M.Kom</u>
NIP. 19890201 201903 1 009 |  |
| 2. Pembimbing
Pendamping | : <u>Mamluatul Hani'ah, S.Kom., M.Kom</u>
NIP. 19900206 201903 2 013 |  |
| 3. Penguji
Utama | : <u>Imam Fahrur Rozi, ST., MT.</u>
NIP. 19840610 200812 1 004 |  |
| 4. Penguji
Pendamping | : <u>Moch. Zawaruddin Abdullah, S.ST., M.Kom</u>
NIP. 19890210 201903 1 019 |  |

Mengetahui,

Ketua Jurusan
Teknologi Informasi



Dr. Eriy Rosa Andrie Asmara, S.T., M.T.
NIP. 19801010 200501 1 001

Ketua Program Studi
Teknik Informatika

Dr. Ely Setyo Astuti, S.T., M.T.
NIP. 19760515 200912 2 001

PERNYATAAN

Dengan ini saya menyatakan bahwa pada Skripsi ini tidak terdapat karya, baik seluruh maupun sebagian, yang sudah pernah diajukan untuk memperoleh gelar akademik di Perguruan Tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini serta disebutkan dalam daftar sitasi/pustaka.

Malang, 26 Juli 2023



Irfan Thalib Alfarid

KATA PENGANTAR

Puji Syukur kami panjatkan kehadirat Allah SWT/Tuhan YME atas segala rahmat dan hidayah-Nya peneliti dapat menyelesaikan skripsi dengan judul “CLUSTERING JAWABAN URAIAN MAHASISWA MENGGUNAKAN TF-IDF DAN KMEANS”. Skripsi ini peneliti susun sebagai persyaratan untuk menyelesaikan studi program Diploma IV Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang.

Kami menyadari bahwasannya dengan tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan laporan akhir ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Bapak Dr. Eng. Rosa Andrie Asmara, S.T., M.T. selaku Ketua Jurusan Teknologi Informasi
2. Ibu Dr. Ely Setyo Astuti, S.T., M.T. selaku Ketua Program Studi DIV Teknik Informatika
3. Bapak Irsyad Arif Mashudi, S.Kom., M.Kom., selaku Pembimbing I yang telah meluangkan waktunya untuk memberikan bimbingan, ilmu, masukan dan dukungan selama penelitian serta pelaksanaan penelitian.
4. Ibu Mamluatul Hani'ah, S.Kom., M.Kom., selaku Pembimbing II yang telah meluangkan waktunya untuk memberikan bimbingan, ilmu, masukan dan dukungan selama penelitian serta pelaksanaan penelitian.
5. Kedua orang tua dan keluarga yang selalu tak henti-hentinya memberikan doa dorongan motivasi dan dorongan moriil maupun materiil.
6. Teman-teman seperjuangan jurusan Teknologi Informasi yang tidak dapat saya sebutkan satu persatu
7. Putri Alivia Nur Kusumawati selaku kekasih yang senantiasa memberikan segala dukungan
8. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Laporan Akhir dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Peneliti menyadari bahwa dalam penyusunan laporan akhir ini, masih banyak terdapat kekurangan dan kelemahan yang dimiliki peneliti baik itu

sistematika penelitian maupun penggunaan bahasa. Untuk itu peneliti mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan laporan ini. Semoga laporan ini berguna bagi pembaca secara umum dan peneliti secara khusus. Akhir kata, peneliti ucapkan banyak terima kasih.

Malang, 26 Juli 2023

Peneliti

ABSTRAK

Alfarid, Irfan Thalib. “*Clustering* Jawaban Uraian Mahasiswa Menggunakan TF-IDF dan *KMeans*”.

Pembimbing: (1) Irsyad Arif Mashudi, S.Kom., M.Kom (2) Mamluatul Hani'ah, S.Kom., M.Kom

Skripsi, Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang, 2023.

Saat ini teknologi sudah menjadi kebutuhan primer manusia yang digunakan di semua segi kehidupan manusia, salah satunya adalah bidang pendidikan. Pada bidang ini penemuan pengetahuan, metode, dan perspektif baru ini tentu tidak bisa ditelan mentah begitu saja, namun diperlukan latihan dan pengarahan yang baik khususnya kepada mahasiswa tujuannya agar pengetahuan tersebut dapat dipahami mahasiswa dengan baik sehingga tidak disalahgunakan. Salah satu cara untuk memastikan mahasiswa memahami betul suatu pengetahuan adalah dengan cara diberi soal uraian, karena soal uraian memberikan evaluasi yang lebih akurat dibandingkan dengan jenis soal lainnya, namun hal ini tentu memunculkan masalah baru dimana dosen seringkali masih belum menemukan cara yang efektif dalam menilai jawaban soal uraian. *Clustering* merupakan salah satu cara yang dapat digunakan untuk mengetahui variasi jawaban siswa secara keseluruhan. TF-IDF dan *KMeans* merupakan algoritma clustering yang dinilai paling kuat dan paling populer. Dengan menggunakan TF-IDF dan *KMeans* dalam membantu dosen mengelompokkan jawaban uraian mahasiswa ternyata cukup efektif karena dengan persentase kesesuaian hasil pengelompokkan sebesar 65% dosen dapat mengelompokkan jawaban uraian dengan waktu jauh lebih cepat dari mengelompokkan jawaban uraian secara manual.

Kata Kunci: Jawaban Soal Uraian, *Clustering*, TF-IDF, *KMeans*

ABSTRACT

Alfarid, Irfan Thalib. “*Student Essay Answer Clustering using TF-IDF and KMeans*”.

Supervisor: (1) Irsyad Arif Mashudi, S.Kom., M.Kom (2) Mamluatul Hani'ah, S.Kom., M.Kom

Thesis, Informatics Engineering Study Program, Information Technology Department, Politeknik Negeri Malang, 2023.

Nowadays, technology has become a primary human need used in all aspects of human life, including education. In this field, the discovery of new knowledge, methods, and perspectives cannot be simply accepted at face value. Instead, it requires proper training and guidance, especially for students, so that this knowledge can be understood and not misused. One way to ensure students really understand a piece of knowledge is by giving them essay questions, because essay questions provide a more accurate evaluation compared to other types of questions. However, this certainly poses new problems where lecturers frequently struggle to find an effective way to assess essay answers. Clustering is a method that can be used to understand variations in student responses as a whole. TF-IDF and KMeans are the most powerful and most popular clustering algorithms. Using TF-IDF and KMeans in helping lecturers to classify student essay answers and has proven to be quite effective. With a matching percentage of 65%, lecturers can categorize essay answers much faster using this method than by doing it manually.

Keywords: *Student Essay Answer, Clustering, TF-IDF, KMeans*

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xv
BAB I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan	3
1.4. Batasan Masalah	3
1.5. Sistematika Penelitian	4
BAB II. LANDASAN TEORI	6
2.1. Dasar Teori	6
2.1.1. Tes Uraian	6
2.1.2. Clustering	6
2.1.3. Teks Preprocessing	7
2.1.4. TF-IDF	8
2.1.5. KMeans	9
2.1.6. Sastrawi	10
2.1.7. Bahasa Pemrograman Python	10
2.1.8. Metode Elbow	10
2.1.9. Metode Silhouette	10
2.1.10. React	11
2.1.11. Flask	11
2.1.12. Use Case Diagram	11

2.1.13. Activity Diagram	11
2.2. Tinjauan Pustaka	12
BAB III. METODOLOGI PENELITIAN	16
3.1. Waktu dan Tempat Penelitian	16
3.2. Metode Penelitian	16
3.2.1. Pengumpulan Data	16
3.2.2. Pengolahan Data	17
3.2.3. Pembuatan Model	20
3.2.4. Penerapan Algoritma	22
3.2.5. Penarikan Kesimpulan	26
BAB IV. ANALISIS DAN PERANCANGAN SISTEM	28
4.1. Deskripsi Sistem	28
4.2. Analisis Permasalahan	28
4.3. Analisis Kebutuhan Sistem	29
4.3.1. Analisis Pengguna	29
4.3.2. Analisis Kebutuhan Fungsional	29
4.3.3. Analisis Kebutuhan Non Fungsional	30
4.4. Perancangan Sistem	31
4.4.1. Perancangan Data	31
4.4.2. Perancangan Algoritma	33
4.4.3. Perancangan Kebutuhan Fungsional	38
4.4.4. Perancangan Antarmuka	51
4.4.5. Perancangan Pengujian	56
BAB V. IMPLEMENTASI DAN PENGUJIAN	69
5.1. Implementasi Data	69
5.1.1. <i>Preprocessing</i>	70
5.2. Implementasi Algoritma	75
5.2.1. <i>Vectorizing</i>	76
5.2.2. <i>Clustering</i>	76
5.2.3. <i>Testing</i>	77
5.3. Implementasi Sistem	78
5.3.1. Implementasi <i>Database</i>	78

5.3.2. Implementasi Antarmuka	87
5.4. Pengujian	106
5.4.1. Pengujian Klaster	106
5.4.2. Pengujian Sistem	107
5.4.3. Pengujian Efektifitas	116
BAB VI. HASIL DAN PEMBAHASAN	118
6.1. Hasil Implementasi	118
6.1.1. Hasil Implementasi Data	118
6.1.2. Hasil Implementasi Algoritma	119
6.1.3. Hasil Data Uji	121
6.2. Pembahasan Data Uji	129
6.2.1. Pembahasan Data Uji Klaster	130
6.2.2. Pembahasan Data Uji Sistem	131
6.2.3. Pembahasan Data Uji Efektifitas	131
BAB VII. KESIMPULAN DAN SARAN	134
7.1. Kesimpulan	134
7.2. Saran	134
DAFTAR PUSTAKA	135
LAMPIRAN	139

DAFTAR GAMBAR

Gambar 3. 1. Tahapan Metode Penelitian	16
Gambar 3. 2. Alur Proses Pengolahan Data	18
Gambar 3. 3. Model Sistem.....	20
Gambar 3. 4. Model Detail Sistem.....	21
Gambar 3. 5. Alur Algoritma <i>TF-IDF</i>	22
Gambar 3. 6. Alur Proses <i>Kmeans Clustering</i>	23
Gambar 3. 7. Alur Proses <i>Silhouette Coefficient</i>	25
Gambar 4. 1. Alur Proses <i>TF-IDF</i>	34
Gambar 4. 2. Alur Proses <i>Kmeans</i>	35
Gambar 4. 3. Alur Proses <i>Silhouette Coefficient</i>	37
Gambar 4. 4. Arsitektur Sistem.....	39
Gambar 4. 5. Use Case Diagram	40
Gambar 4. 6. Activity Diagram Login	41
Gambar 4. 7. Activity Diagram Kelola User.....	42
Gambar 4. 8. Activity Diagram Kelola Kategori	43
Gambar 4. 9. Activity Diagram Kelola Soal	44
Gambar 4. 10. Activity Diagram Lihat Klaster.....	45
Gambar 4. 11. Activity Diagram Jawab Soal.....	46
Gambar 4. 12. Perancangan Relasi Tabel	51
Gambar 4. 13. Perancangan Antarmuka Login.....	52
Gambar 4. 14. Perancangan Antarmuka Kelola User	53
Gambar 4. 15. Perancangan Antarmuka Kelola Kategori.....	53
Gambar 4. 16. Perancangan Antarmuka Kelola Soal.....	54
Gambar 4. 17. Perancangan Antarmuka Lihat Klaster	55
Gambar 4. 18. Perancangan Antarmuka Jawab Soal	55
Gambar 4. 19. Proses <i>Silhouette Coefficient</i>	57
Gambar 4. 20. Use Case Diagram	59
Gambar 5. 1. Implementasi Relasi Tabel	86

Gambar 5. 2. Halaman Antarmuka Login.....	87
Gambar 5. 3. Halaman Antarmuka Dasbor Admin.....	88
Gambar 5. 4. Halaman Antarmuka Daftar User.....	89
Gambar 5. 5. Halaman Antarmuka Form Tambah User Baru	90
Gambar 5. 6. Halaman Antarmuka Form Edit User.....	91
Gambar 5. 7. Halaman Antarmuka Daftar Kategori	92
Gambar 5. 8. Halaman Antarmuka Tambah Kategori Baru.....	93
Gambar 5. 9. Halaman Antarmuka Edit Kategori.....	94
Gambar 5. 10. Halaman Antarmuka Dasbor Dosen.....	95
Gambar 5. 11. Halaman Antarmuka Daftar Soal	96
Gambar 5. 12. Halaman Antarmuka Tambah Soal Baru	97
Gambar 5. 13. Halaman Antarmuka Form Edit Soal	98
Gambar 5. 14. Halaman Antarmuka Daftar Jawaban	99
Gambar 5. 15. Halaman Antarmuka Pilih Soal.....	100
Gambar 5. 16. Halaman Antarmuka Lihat Klaster	102
Gambar 5. 17. Halaman Antarmuka Dasbor Mahasiswa.....	104
Gambar 5. 18. Halaman Antarmuka Pilih Kategori Soal.....	104
Gambar 5. 19. Halaman Antarmuka Jawab Soal	105
Gambar 5. 20. Use Case Diagram	108
Gambar 6. 1. Grafik Pesebaran Nilai <i>Silhouette Coefficient</i> Pada Klaster	130

DAFTAR TABEL

Tabel 2. 1. Penelitian Terdahulu	12
Tabel 3. 1. Contoh Data Soal dan Jawaban.....	17
Tabel 3. 2. Contoh Proses <i>Case Folding</i>	18
Tabel 3. 3. Contoh Proses <i>Tokenization</i>	18
Tabel 3. 4. Contoh Proses <i>Filtering & Stemming</i>	19
Tabel 3. 5. Contoh Proses <i>Stopword Removal</i>	19
Tabel 3. 6. Contoh Proses <i>Preprocessing</i>	19
Tabel 3. 7. Contoh Pengujian Efektifitas	26
Tabel 4. 1. Kebutuhan Fungsional Sistem	29
Tabel 4. 2. Kebutuhan Perangkat Lunak	30
Tabel 4. 3. Kebutuhan Perangkat Keras.....	30
Tabel 4. 4. Contoh Data Kotor	31
Tabel 4. 5. Contoh Hasil <i>Case Folding</i>	31
Tabel 4. 6. Contoh Hasil <i>Tokenization</i>	32
Tabel 4. 7. Contoh Hasil <i>Stemming</i>	32
Tabel 4. 8. Contoh Hasil <i>Stopword Removal</i>	33
Tabel 4. 9. Contoh Perbandingan Data Kotor dan Data Bersih	33
Tabel 4. 10. Rancangan Tabel Users.....	47
Tabel 4. 11. Rancangan Tabel Kategoris	47
Tabel 4. 12. Rancangan Tabel Soals	48
Tabel 4. 13. Rancangan Tabel Jawabs	49
Tabel 4. 14. Rancangan Tabel Sessions	50
Tabel 4. 15. Skenario Login.....	59
Tabel 4. 16 Skenario Kelola User	60
Tabel 4. 17. Skenario Kelola Kategori.....	62
Tabel 4. 18. Skenario Kelola Soal.....	63
Tabel 4. 19. Skenario Lihat Klaster	64
Tabel 4. 20. Skenario Jawab Soal	65

Tabel 4. 21 Skenario Logout.....	66
Tabel 4. 22. Perancangan Pengujian Efektifitas	67
Tabel 5. 1. Sampel Data	69
Tabel 5. 2. Sampel Data <i>Case Folding</i>	70
Tabel 5. 3. Sampel Data <i>Tokenization</i>	71
Tabel 5. 4. Sampel Data <i>Stemming</i>	72
Tabel 5. 5. Sampel Data <i>Stopword Removal</i>	74
Tabel 5. 6. Implementasi Tabel Users.....	78
Tabel 5. 7. Implementasi Tabel Kategoris	80
Tabel 5. 8. Implementasi Tabel Soals	81
Tabel 5. 9. Implementasi Tabel Jawabs	83
Tabel 5. 10. Implementasi Tabel Sessions	85
Tabel 5. 11. Pengujian Proses Login.....	109
Tabel 5. 12. Pengujian Proses Kelola User	109
Tabel 5. 13. Pengujian Proses Kelola Kategori	111
Tabel 5. 14. Pengujian Proses Kelola Soal	112
Tabel 5. 15. Pengujian Proses Lihat Klaster	113
Tabel 5. 16. Pengujian Proses Jawab Soal	114
Tabel 5. 17. Pengujian Proses Logout.....	115
Tabel 5. 18. Pengujian Efektifitas	116
Tabel 6. 1. Sampel Hasil Implementasi Data.....	118
Tabel 6. 2. Hasil Implementasi Algoritma <i>Vectorizing</i>	119
Tabel 6. 3. Hasil Implementasi Algoritma <i>Clustering</i>	120
Tabel 6. 4. Hasil Implementasi Algoritma <i>Silhouette Coefficient</i>	120
Tabel 6. 5. Hasil Implementasi Pengujian Klaster	121
Tabel 6. 6. Hasil Implementasi Pengujian Sistem	122
Tabel 6. 7. Hasil Implementasi Pengujian Efektifitas.....	123
Tabel 6. 8. Persentase Hasil Implementasi Pengujian Sistem.....	131
Tabel 6. 9. Perbandingan Hasil Implementasi Pengujian Efektifitas.....	132

DAFTAR LAMPIRAN

Lampiran 1. Hasil <i>Preprocessing</i> Jawaban Uraian	139
Lampiran 2. Hasil <i>Vectorizing</i> Jawaban Uraian.....	154
Lampiran 3. Hasil <i>Clustering</i> dan <i>Testing</i> Jawaban Uraian.....	155
Lampiran 4. Grafik Hasil <i>Testing</i> Kluster	156