

SKRIPSI

**ANALISIS ALGORITMA CLUSTER CONDITIONAL
MARKOV RANDOM WALK DENGAN METODE
EKSTRAKTIF DAN VARIASI FREKUENSI TF-IDF PADA
PERINGKASAN MULTI-DOCUMENT BERBAHASA
INDONESIA**



WAHID RIZQI AKBAR NURJAMAN

5235164349

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMASI DAN KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS NEGERI

JAKARTA

2023

SKRIPSI

**ANALISIS ALGORITMA CLUSTER CONDITIONAL
MARKOV RANDOM WALK DENGAN METODE
EKSTRAKTIF DAN VARIASI FREKUENSI TF-IDF PADA
PERINGKASAN MULTI-DOCUMENT BERBAHASA
INDONESIA**



WAHID RIZQI AKBAR NURJAMAN

5235164349

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMASI DAN KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS NEGERI

JAKARTA

2023



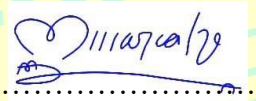
HALAMAN PENGESAHAN

**Analisis Algoritma Cluster Conditional Markov Random Walk dengan
Metode Ekstraktif dan Variasi TF-IDF pada Peringkasan *Multi-document*
berbahasa Indonesia**

Wahid Rizqi Akbar Nurjaman, NIM. 5235164349

NAMA DOSEN	TANDA TANGAN	TANGGAL
Dr. Widodo, M.Kom. Dosen Pembimbing I		3 April 2023
Irma Permata Sari, S.Pd., M.Eng. Dosen Pembimbing II		3 April 2023

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Hamidillah Ajie, S.Si, M.T (Ketua Penguji)		18 Mei 2023
Diat Nurhidayat, S.Pd, M.TI (Dosen Penguji I)		18 Mei 2023
Murien Nugraheni, S.Pd, M.Cs. (Dosen Penguji II)		16 Mei 2023

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 18 Mei 2023

Yang membuat pernyataan



Wahid Rizqi Akbar Nurjaman
NIM. 5245164349



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Wahid Rizqi Akbar Nurjaman
NIM : 5235164349
Fakultas/Prodi : Teknik / Pendidikan Teknik Informatika dan Komputer
Alamat email : wahidrizqiakbar@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Analisis Algoritma Cluster Conditional Markov Random Walk dengan Metode Ekstraktif dan Variasi TF-IDF pada Peringkasan Multi-document berbahasa Indonesia

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta 1 September 2023

Penulis

(Wahid Rizqi Akbar N.)
nama dan tanda tangan

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul “**Analisis Algoritma Cluster Conditional Markov Random Walk dengan Metode Ekstraktif dan Variasi TF-IDF pada Peringkasan *Multi-document* berbahasa Indonesia**”, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Program Studi Pendidikan Teknik Infromasi dan Komputer.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih yang setulus-tulusnya kepada:

1. Bapak Dr. Widodo, M.Kom, selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta sekaligus Dosen Pembimbing I yang telah memberikan bimbingan, arahan serta saran yang diberikan kepada penulis sehingga skripsi ini dapat terselesaikan dengan baik;
2. Ibu Irma Permata Sari, S.Pd., M.Eng, selaku dosen pembimbing II skripsi atas segala bimbingan, arahan serta saran yang diberikan kepada penulis sehingga skripsi ini dapat terselesaikan dengan baik;
3. Seluruh Bapak dan Ibu Dosen Pendidikan Teknik Informatika dan Komputer Universitas Negeri Jakarta yang telah memberikan ilmu pengetahuan yang tak ternilai selama penulis menempuh pendidikan;
4. Ayahanda penulis, Jaja Nurjaman serta adik saya Afifah Dinah Zhafirah yang selalu memberikan kasih sayang, doa, nasehat, serta atas kesabarannya yang luar biasa dalam setiap langkah hidup penulis, yang merupakan anugerah terbesar dalam hidup. Penulis berharap dapat menjadi anak yang dapat dibanggakan;
5. Teruntuk Mahvira Fitra Aprilia, terima kasih banyak telah mensupport dan percaya bahwa penulis dapat menyelesaikan proses studi, terima kasih atas

doa, cinta, perhatian, dan kasih sayang yang selalu diberikan sehingga penulis bisa menyelesaikan tugas akhir dengan sebaik-baiknya, walau prosesnya tidak mudah penuh rintangan, pengorbanan, suka dan duka, tapi kamu tidak pernah menyerah untuk selalu menjadi support system yang terbaik. Terima kasih cintaku;

6. Teman-teman Alikos. Deki, Yoga, Dipa, Aul, Swardi, Calvin. Terima kasih atas kesenangan, canda tawa yang membahagiakan dan menjadi keluarga baru bagi penulis;
7. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah dengan tulus ikhlas membantu, memberikan doa dan motivasi sehingga dapat terselesaikannya skripsi ini.

Dalam penulisan skripsi ini masih banyak kekurangan dan kesalahan, karena itu segala kritik dan saran yang membangun akan menyempurnakan penulisan skripsi ini serta bermanfaat bagi penulis dan pembaca.

Jakarta, 18 Mei 2023

Penulis,

Wahid Rizqi Akbar Nurjaman

5245164349

ABSTRAK

Informasi merupakan salah satu bagian terpenting dalam sejarah kehidupan manusia. Dalam ranah ilmu pengetahuan, informasi yang berharga dan dapat dipertanggungjawabkan kebenarannya ditulis dalam bentuk jurnal ilmiah, artikel atau dokumen. Semakin terbiasanya masyarakat Indonesia menggunakan teknologi komputer maupun *smartphone* dan banyaknya informasi yang beredar di *Internet*, berhasil menyediakan kemudahan untuk mendapatkan pengetahuan yang mereka inginkan dengan sentuhan jari. Salah satu metode untuk memudahkan temu kembali informasi, diperlukan adanya peringkasan teks terlebih dahulu dan pengetahuan yang telah berkembang menghasilkan peringkasan teks otomatis dengan model tertentu. Penelitian ini bertujuan untuk menganalisa model algoritma *Conditional Cluster-based Markov Random Walk* dengan sudah diterapkan preprocessing dan teknik *clustering K-Means* untuk peringkasan teks otomatis. *Dataset* yang digunakan sudah disetujui oleh pakar Bahasa Indonesia dengan topik utama *Natural Language Processing*. Hasil model ini, dengan analisa ROUGE-1 menunjukkan performa terbaik untuk variasi TF-IDF *standard* dengan nilai presisi 0.495, *recall* 0.294 dan *F-measure* 0.370 pada 10 dokumen dan hasil ringkasan 20 kalimat. Sementara untuk hasil variasi TF-IDF dengan filter *thesaurus* Sinonim menunjukkan performa terbaik dengan nilai persisi 0.543, *recall* 0.282 dan *F-measure* 0.372 pada 10 dokumen dan hasil ringkasan 20 kalimat.

Kata kunci: peringkasan teks otomatis, k-means, metode ekstraktif, markov random walk, ROUGE-1, bahasa Indonesia

ABSTRACT

Information is one of the most important parts of the history of human life. In the realm of science, valuable and reliable information is written in the form of scientific journals, articles, or documents. The more accustomed Indonesian people are to using computer and smartphone technology and the amount of information circulating on the Internet, successfully providing the convenience of getting the knowledge they want at the touch of a finger. One of the methods to facilitate information retrieval, text summarization is needed first, and the knowledge that has developed results in automatic text summarization with certain models. This research aims to analyze the Conditional Cluster-based Markov Random Walk algorithm model with preprocessing and K-Means clustering techniques for automatic text summarization. The dataset used has been approved by Indonesian language experts with the main topic of Natural Language Processing. The results of this model, with ROUGE-1 analysis, show the best performance for standard TF-IDF variation with precision value 0.495, recall 0.294, and F-measure 0.370 on 10 documents and 20 sentences summary results. While for the result of the TF-IDF variation with the Synonym filter, shows the best performance with a precision value of 0.543, recall of 0.282, and F-measure 0.372 on 10 documents and 20 sentences summary results.

Keywords: automatic text summarization, k-means, extractive method, Markov random walk, ROUGE-1, Indonesian language

DAFTAR ISI

LEMBAR PERSETUJUAN JUDUL SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
KATA PENGANTAR	v
ABSTRAK	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Identifikasi Masalah	3
1.3. Pembatasan Masalah	3
1.4. Rumusan Masalah	4
1.5. Tujuan Penelitian.....	4
1.6. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kerangka Teoritik.....	5
2.1.1 Peringkasan Teks	5
2.1.2 Peringkasan Teks Otomatis Multi Dokumen	6
2.1.3 Peringkasan Teks Otomatis dengan Pendekatan Ekstraktif	6
2.1.4 Term Frequency - Inverse Document Frequency (TF-IDF)	7
2.1.5 K-Means	8
2.1.6 Cluster-Based Conditional Markov Random Walk Model.....	8
2.2 Penelitian Relevan	10
2.3 Kerangka Berpikir	14
BAB III METODOLOGI PENELITIAN	16
3.1. Tempat dan Waktu Penelitian	16
3.2. Alat dan Bahan Penelitian	16
3.2.1 Alat Penelitian.....	16
3.2.2 Bahan Penelitian.....	17
3.3. Diagram Alur Penelitian.....	17
3.3.1. Pengumpulan Data	17
3.3.2. Preprocessing	18

3.3.3.	Pembobotan Kata dengan TF-IDF	19
3.3.4.	Penghilangan Kata Sinonim dengan Tesaurus Bahasa Indonesia ...	20
3.3.5.	Clustering Kalimat dengan K-Means	21
3.3.6.	Saliency Score dengan Conditional Markov Random Walk.....	21
3.4.	Teknik dan Prosedur Pengumpulan Data	21
3.4.1	Teknik Pengumpulan Data	21
3.4.2	Prosedur Pengumpulan Data	22
3.5.	Teknik Analisis Data	22
BAB IV	HASIL PENELITIAN	24
4.1.	Deskripsi Hasil Penelitian	24
4.2.	Analisis Data Penelitian	54
4.3.	Pembahasan	55
4.4.	Aplikasi Hasil Penelitian	56
BAB V	KESIMPULAN DAN SARAN	57
5.1.	Kesimpulan.....	57
5.2.	Saran.....	57
DAFTAR PUSTAKA	59

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2.1	Penelitian Relevan	13
3.1	Daftar Spesifikasi Perangkat Keras	16
3.2	Daftar Spesifikasi Perangkat Lunak	16



DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
2.1	Dua Lapis Grafik Tautan (Wan dan Yang, 2008)	19
2.2	Bagan Kerangka Berpikir	26
3.1	Diagram Alur Penelitian	28



DAFTAR LAMPIRAN

Nomor	Judul Lampiran	Halaman
1	Surat Tugas Dosen Pembimbing Skripsi	61
2	Lembar Bimbingan Dosen Pembimbing I	62
3	Lembar Bimbingan Dosen Pembimbing II	63
4	Surat Pernyataan Persetujuan Pendaftaran Sidang Skripsi Dosen Pembimbing I	64
5	Surat Pernyataan Persetujuan Pendaftaran Sidang Skripsi Dosen Pembimbing II	65
6	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (10 Dokumen, 5 Kalimat)	66
7	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (10 Dokumen, 10 Kalimat)	67
8	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (10 Dokumen, 20 Kalimat)	68
9	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (20 Dokumen, 5 Kalimat)	69
10	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (20 Dokumen, 10 Kalimat)	70
11	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (20 Dokumen, 20 Kalimat)	71
12	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (50 Dokumen, 5 Kalimat)	72
13	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (50 Dokumen, 10 Kalimat)	73
14	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Berdasarkan Pakar (50 Dokumen, 20 Kalimat)	74
15	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (10 Dokumen, 5 Kalimat)	76

16	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (10 Dokumen, 10 Kalimat)	77
17	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (10 Dokumen, 20 Kalimat)	78
18	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (20 Dokumen, 5 Kalimat)	79
19	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (20 Dokumen, 10 Kalimat)	80
20	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (20 Dokumen, 20 Kalimat)	81
21	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (50 Dokumen, 5 Kalimat)	82
22	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (50 Dokumen, 10 Kalimat)	83
23	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 1 (50 Dokumen, 20 Kalimat)	84
24	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (10 Dokumen, 5 Kalimat)	85
25	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (10 Dokumen, 10 Kalimat)	86

26	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (10 Dokumen, 20 Kalimat)	87
27	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (20 Dokumen, 5 Kalimat)	88
28	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (20 Dokumen, 10 Kalimat)	89
29	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (20 Dokumen, 20 Kalimat)	90
30	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (50 Dokumen, 5 Kalimat)	91
31	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (50 Dokumen, 10 Kalimat)	92
32	Hasil Peringkasan Multi Dokumen Ilmiah Berbahasa Indonesia Menggunakan Model CCMRW Varian 2 (50 Dokumen, 20 Kalimat)	93