

**PENERAPAN METODE *MAXIMUM MARGINAL RELEVANCE*
DALAM PERINGKASAN TEKS OTOMATIS PADA PORTAL
BERITA BERBAHASA INDONESIA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

GUSTIO NUSAMBA

19.11.3179

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**PENERAPAN METODE *MAXIMUM MARGINAL RELEVANCE*
DALAM PERINGKASAN TEKS OTOMATIS PADA PORTAL
BERITA BERBAHASA INDONESIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

GUSTIO NUSAMBA

19.11.3179

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN METODE *MAXIMUM MARGINAL RELEVANCE* DALAM
PERINGKASAN TEKS OTOMATIS PADA PORTAL BERITA**

BERBAHASA INDONESIA

yang disusun dan diajukan oleh

Gustio Nusamba

19.11.3179

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Februari 2023

Dosen Pembimbing,



Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108

HALAMAN PENGESAHAN
SKRIPSI
PENERAPAN METODE *MAXIMUM MARGINAL RELEVANCE* DALAM
PERINGKASAN TEKS OTOMATIS PADA PORTAL BERITA
BERBAHASA INDONESIA

yang disusun dan diajukan oleh

Gustio Nusamba

19.11.3179

Telah dipertahankan di depan Dewan Penguji

pada tanggal 22 Februari 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, M.Kom

NIK. 190302185

Theopilus Bayu Sasongko, S.Kom, M.Eng

NIK. 190302375

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Februari 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Gustio Nusamba
NIM : 19.11.3179

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Metode Maximum Marginal Relevance Dalam Peringkasan Teks Otomatis Pada Portal Berita Berbahasa Indonesia

Dosen Pembimbing : Mardhiya Hayaty, S. T., M. Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Februari 2023

Yang Menyatakan,

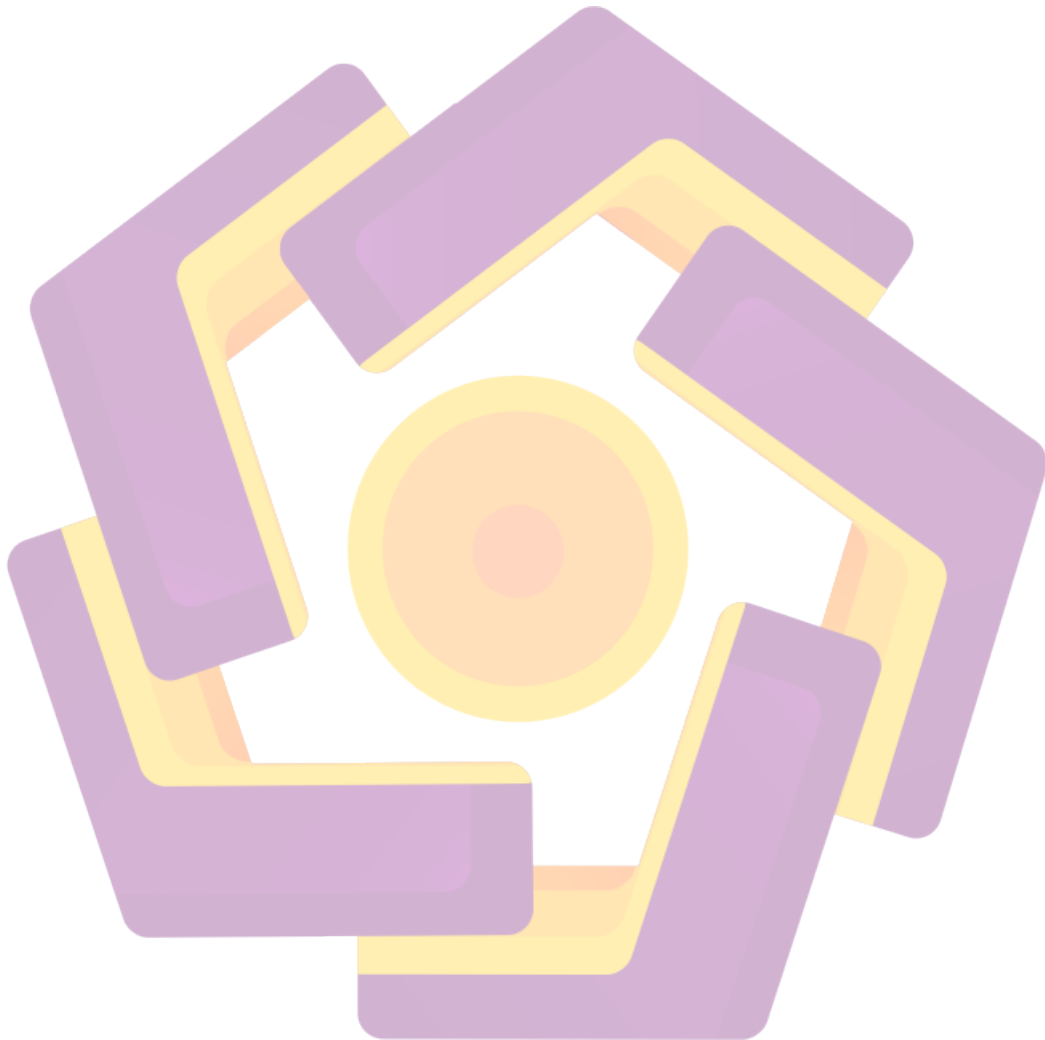


Gustio Nusamba

MOTTO

"After climbing a great hill, one only finds that there are many more hills to climb"

(Nelson Mandela)



PERSEMBAHAN

Dengan penuh rasa syukur atas kesempatan yang telah Allah SWT berikan dan dengan telah diselesaikannya skripsi ini. Saya mempersembahkan skripsi ini kepada :

1. Bapak dan Ibu saya tercinta yang telah memberi dukungan moral, mendoakan, dan selalu memperhatikan kesehatan saya di setiap kondisi.
2. Adik saya, Hatmaja Narotama yang mengajak saya untuk *push rank game Mobile Legends* untuk mengisi waktu di saat saya mulai pusing mengerjakan skripsi.
3. Ibu Mardhiya Hayaty, S. T., M. Kom, selaku dosen pembimbing yang selalu membimbing, mengarahkan, dan bahkan memudahkan jalannya proses skripsi ini.
4. Teman saya Abi Mas Hardianto yang membantu saya memberikan solusi dalam proses pengerjaan skripsi dan juga sebagai tempat saya untuk berdiskusi.
5. Teman kelas saya yaitu kelas IF-09 yang selalu menemani saya ketika saya merasa bosan dan mengajak saya untuk pergi *healing*.
6. Serta semua pihak yang membantu dan mendukung saya yang tidak bisa saya sebutkan satu persatu.

KATA PENGANTAR

Puji dan syukur saya panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunianya sehingga saya dapat menyelesaikan usulan penelitian yang berjudul “Penerapan Metode Maximum Marginal Relevance Dalam Peringkasan Teks Otomatis Pada Portal Berita Berbahasa Indonesia”.

Skripsi ini membahas tentang hasil implementasi sebuah metode *Maximum Marginal Relevance* pada peringkasan teks berita berbahasa Indonesia. Kajian tersebut didasarkan pada banyaknya informasi saat ini yang tersebar di berbagai situs berita *online* yang menuntut para pembaca untuk cepat dalam mendapatkan informasi yang tersirat.

Proses penyusunan hingga selesainya laporan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Maka dari itu, saya tidak lupa mengucapkan terima kasih dan penghargaan kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S. Kom., M. Kom, selaku Dekan Fakultas Ilmu Universitas AMIKOM Yogyakarta.
3. Ibu Mardhiya Hayaty, S.T., M. Kom, selaku dosen pembimbing saya yang telah memberikan bimbingan dan *support* penuh dalam proses penulisan skripsi ini.
4. Bapak, Ibu, dan Adik saya tercinta yang telah memberikan kasih sayang, motivasi, doa, dan dukungan terhadap saya.
5. Bapak dan Ibu Dosen serta staf Universitas Amikom Yogyakarta yang telah banyak memberikan ilmunya selama saya berkuliah.
6. Teman-teman 19-IF-09 atas segala bantuan, dukungan, dan doanya.
7. Semua pihak yang telah membantu dalam penyelesaian skripsi ini yang tidak dapat saya sebutkan satu persatu
8. Akhir kata, kritik dan masukan dari semua pihak sangat saya harapkan.

Saya menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Sekecil apapun saya berharap, skripsi ini dapat memberikan kontribusi positif terhadap upaya-upaya nyata dari digitalisasi kehidupan tentunya untuk memudahkan pembaca agar memperoleh informasi secara efektif dan efisien. Apabila terdapat kesalahan dan kekurangan mohon maaf sebesar-besarnya.

Yogyakarta, 26 Januari 2023

Gustio Nusamba



DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.2 Dasar Teori	11
2.2.1 Automatic Text Summarization (Peringkasan Teks Otomatis)	11
2.2.2 Ringkasan.....	11
2.3 Machine Learning	12
2.4 Text Preprocessing.....	12
2.4.1 Pemecahan Kalimat.....	13
2.4.2 Case Folding	13
2.4.3 Tokenizing	13
2.4.4 Filtering.....	14
2.4.5 Stemming	14
2.5 Term Frequency - Inverse Document Frequency (TF-IDF)	17
2.6 Cosine Similarity	19
2.7 Maximum Marginal Relevance (MMR)	20
2.8 Confusion Matrix	22
BAB III METODE PENELITIAN	24

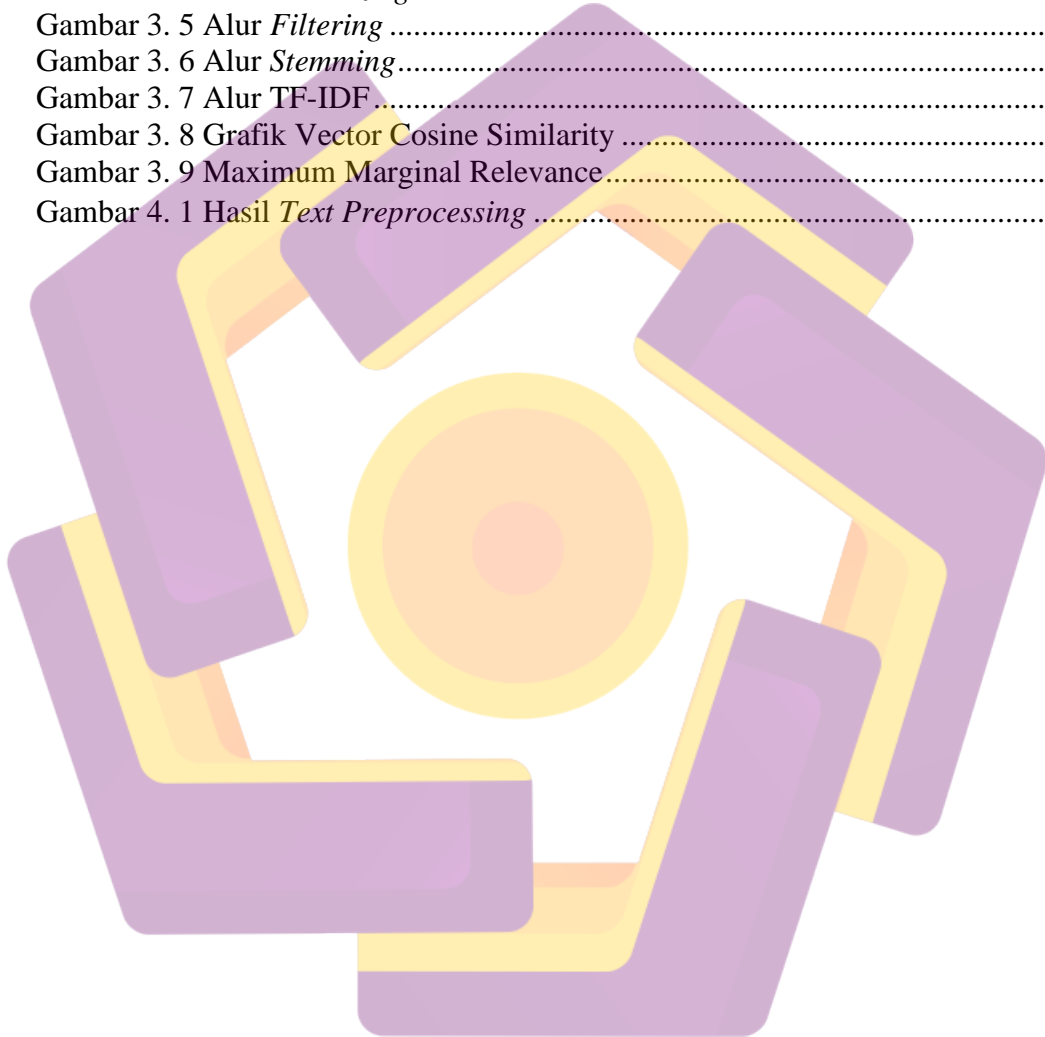
3.1	Alat dan Bahan Penelitian.....	24
3.1.1	Alat Penelitian.....	24
3.1.2	Bahan Penelitian	24
3.2	Alur Penelitian	28
3.3	Metode Pengolahan Data	30
3.3.1	Text Preprocessing.....	30
3.3.1.1	Sentence Segmentation.....	31
3.3.1.2	Case Folding	32
3.3.1.3	Tokenizing	32
3.3.1.4	Filtering.....	34
3.3.1.5	Stemming.....	35
3.3.1	Term Frequency – Inverse Document Frequency (TF-IDF).....	35
3.3.2	Cosine Similarity.....	37
3.3.3	Maximum Marginal Relevance.....	38
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Input Dataset	40
4.2	Text Preprocessing.....	40
4.2.1	Sentence Segmentation	41
4.2.2	Case Folding	41
4.2.3	Tokenizing	42
4.2.4	Filtering.....	44
4.2.5	Stemming	46
4.3	Perhitungan TF-IDF (Term Frequency – Inverse Document Frequency) 48	
4.4	Cosine Similarity Antar Kalimat	54
4.5	Cosine Similarity Query Terhadap Kalimat	58
4.6	Metode MMR	62
4.7	Hasil Ringkasan	68
4.8	Analisis Pengujian	69
4.9	Hasil Pengujian Performa Nilai	72
BAB V PENUTUP.....		77
5.1	Kesimpulan	77
5.2	Saran	77
DAFTAR PUSTAKA		78

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	6
Tabel 2. 2 Aturan Pemenggalan Stemmer Nazief dan Andriani.....	16
Tabel 3. 1 Alat Penelitian.....	24
Tabel 3. 2 Dataset Berita Nasional Indonesia.....	25
Tabel 4. 1 Input Dataset.....	40
Tabel 4. 2 Hasil <i>sentence segmentation</i>	41
Tabel 4. 3 Hasil <i>case folding</i>	42
Tabel 4. 4 Hasil <i>tokenizing</i>	42
Tabel 4. 5 Hasil <i>filtering</i>	44
Tabel 4. 6 <i>Stemming</i>	46
Tabel 4. 7 Hasil <i>Stemming</i> menjadi kalimat.....	48
Tabel 4. 8 Hasil Perhitungan TF.....	49
Tabel 4. 9 Hasil perhitungan DF dan IDF.....	51
Tabel 4. 10 Hasil Pembobotan TF-IDF.....	53
Tabel 4. 11 Perhitungan \sum TF-IDF2 D1.....	55
Tabel 4. 12 Perhitungan panjang vektor.....	55
Tabel 4. 13 Hasil pembobotan TF-IDF D1 sampai D8.....	56
Tabel 4. 14 Hasil perkalian antara D1 dengan dokumen lain ($D1 * Di$).....	56
Tabel 4. 15 Perhitungan <i>Cosine Similarity</i> Dokumen 1 (D1).....	57
Tabel 4. 16 Hasil <i>Cosine Similarity</i> Antar Dokumen.....	58
Tabel 4. 17 Hasil total <i>cosine similarity</i> dokumen satu dengan dokumen lainnya.....	58
Tabel 4. 18 Perhitungan \sum TF-IDF2 <i>Query</i>	60
Tabel 4. 19 Hasil perkalian <i>query</i> dengan dokumen lainnya.....	60
Tabel 4. 20 Hasil perkalian antara <i>query</i> dengan dokumen lain ($Q * Di$).....	61
Tabel 4. 21 Perhitungan <i>Cosine Similarity</i> antara <i>query</i> dengan dokumen lain.....	61
Tabel 4. 22 Perhitungan MMR Iterasi ke-1.....	63
Tabel 4. 23 Perhitungan MMR Iterasi ke-2.....	64
Tabel 4. 24 Perhitungan MMR Iterasi ke-3.....	64
Tabel 4. 25 Perhitungan MMR Iterasi ke-4.....	65
Tabel 4. 26 Perhitungan MMR Iterasi ke-5.....	65
Tabel 4. 27 Perhitungan MMR Iterasi ke-6.....	66
Tabel 4. 28 <i>Cosine similarity</i> S1, S2, S3, S5, S8.....	66
Tabel 4. 29 Hasil perankingan kalimat.....	68
Tabel 4. 30 Hasil Ringkasan.....	69
Tabel 4. 31 Perbandingan Hasil Ringkasan.....	71
Tabel 4. 32 Hasil Pengujian Performansi Metode MMR.....	72

DAFTAR GAMBAR

Gambar 2. 1 Tahapan <i>Text Preprocessing</i>	13
Gambar 3. 1 Alur Penelitian	29
Gambar 3. 2 Alur Segmentasi Kalimat	31
Gambar 3. 3 Alur <i>Case Folding</i>	32
Gambar 3. 4 Alur <i>Tokenizing</i>	33
Gambar 3. 5 Alur <i>Filtering</i>	34
Gambar 3. 6 Alur <i>Stemming</i>	35
Gambar 3. 7 Alur TF-IDF	37
Gambar 3. 8 Grafik Vector Cosine Similarity	37
Gambar 3. 9 Maximum Marginal Relevance	38
Gambar 4. 1 Hasil <i>Text Preprocessing</i>	41



INTISARI

Berita adalah kumpulan informasi yang menyampaikan peristiwa yang sedang terjadi. Berita memuat fakta atau opini yang membuat tertarik untuk mengetahuinya. Berita dapat diperoleh dari berbagai media seperti koran, surat kabar, internet dan lain-lain. Seiring perkembangan teknologi berdampak pada bertambahnya jumlah portal berita berbahasa Indonesia. Hal ini menuntut semua informasi bisa diakses dengan cepat dan tidak membutuhkan waktu membaca yang lama.

Demi meminimalkan waktu dalam membaca, maka perlu dilakukan ringkasan teks berita secara otomatis. Teknologi peringkasan teks otomatis menawarkan solusi untuk membantu memperoleh informasi dari suatu berita secara efisien dan efektif. Penelitian diawali dengan lima tahap *text preprocessing* yaitu pemecahan kalimat, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses selanjutnya adalah menghitung bobot TF-IDF, bobot *query relevance*, dan bobot *similarity*. Ringkasan yang dihasilkan diperoleh dari *extractive summarization* yang digunakan untuk meringkas *single document*. Metode *extractive summarization* yang digunakan dalam penelitian ini adalah *Maximum Marginal Relevance* (MMR). MMR meringkas dokumen dengan menghitung kesamaan (*similarity*) antara kalimat dengan kalimat lain dan antara query dengan kalimat.

Dokumen berita yang digunakan dalam penelitian ini diambil 35 berita dari 4 situs berita *online* yaitu *viva.co.id*, *detik.com*, *merdeka.com*, dan *kompas.com*. Hasil pengujian dari 35 berita menggunakan metode MMR dibandingkan dengan ringkasan manual diperoleh rata-rata *recall* 71%, *precision* 79%, *f-measure* 73%, dan akurasi 81%.

Kata kunci: MMR, Peringkasan Teks Otomatis, TF-IDF, Cosine Similarity, Extractive Summarization

ABSTRACT

News is a collection of information that conveys events that are currently happening. News contains facts or opinions that make it interesting to know. News can be obtained from various media such as newspapers, newspapers, internet and others. As technological developments have an impact on the increasing number of Indonesian-language news portals. This requires that all information can be accessed quickly and does not require a long reading time.

In order to minimize reading time, it is necessary to do a news text summary automatically. Automatic text shortening technology offers a solution to help obtain information from a story efficiently and effectively. The research begins with five stages of text preprocessing, namely sentence solving, case folding, tokenizing, filtering, and stemming. The next process is to calculate the TF-IDF weights, query relevance weights, and similarity weights. The resulting summary is obtained from extractive summarization which is used to summarize a single document. The extractive summarization method used in this study is Maximum Marginal Relevance (MMR). MMR summarizes documents by calculating similarities between sentences and other sentences and between queries and sentences.

The news documents used in this study were taken from 35 news sites from 4 online news sites namely viva.co.id, detik.com, merdeka.com, and kompas.com. The test results of 35 news using the MMR method compared to manual summaries obtained an average recall of 71%, precision of 79%, f-measure of 73%, and accuracy of 81%.

Keyword: *MMR, Automatic Text Summarization, TF-IDF, Cosine Similarity, Extractive Summarization*