

IDENTIFIKASI TOPIC MODELLING PADA ULASAN PRODUK

DENGAN ANALISIS LATENT SEMANTIC SIMILARITY

SKRIPSI



disusun oleh

Aulia Rahma Damayanti

17.11.1049

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

UNIVERSITAS AMIKOM YOGYAKARTA

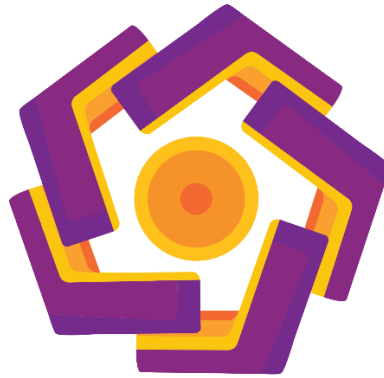
YOGYAKARTA

2022

**IDENTIFIKASI TOPIC MODELLING PADA ULASAN PRODUK
DENGAN ANALISIS LATENT SEMANTIC SIMILARITY**

SKRIPSI

“ Untuk memenuhi sebagian persyaratan mencapai gelar Sarjana pada Program Studi
Informatika”



disusun oleh

Aulia Rahma Damayanti

17.11.1049

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

PERSETUJUAN

SKRIPSI

**IDENTIFIKASI TOPIC MODELLING PADA ULASAN PRODUK
DENGAN ANALISIS LATENT SEMANTIC SIMILARITY**

yang dipersiapkan dan disusun oleh

Aulia Rahma Damayanti

17.11.1049

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 18 November 2021

Dosen Pembimbing,



Agit Amrullah, S. Kom., M. Kom

NIK. 190302356

PENGESAHAN
SKRIPSI
IDENTIFIKASI TOPIC MODELLING PADA ULASAN PRODUK
DENGAN ANALISIS LATENT SEMANTIC SIMILARITY

yang dipersiapkan dan disusun oleh

Aulia Rahma Damayanti

17.11.1049

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Maret 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ali Mustopa, M. Kom

NIK. 190302192

Andi Sunyoto, M. Kom., Dr.

NIK. 190302052

Agit Amrullah, S. Kom., M. Kom.

NIK. 190302356



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 25 Maret 2022

DEKAN FAKULTAS ILMU KOMPUTER

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

NIK. 190302096

PERNYATAAN

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

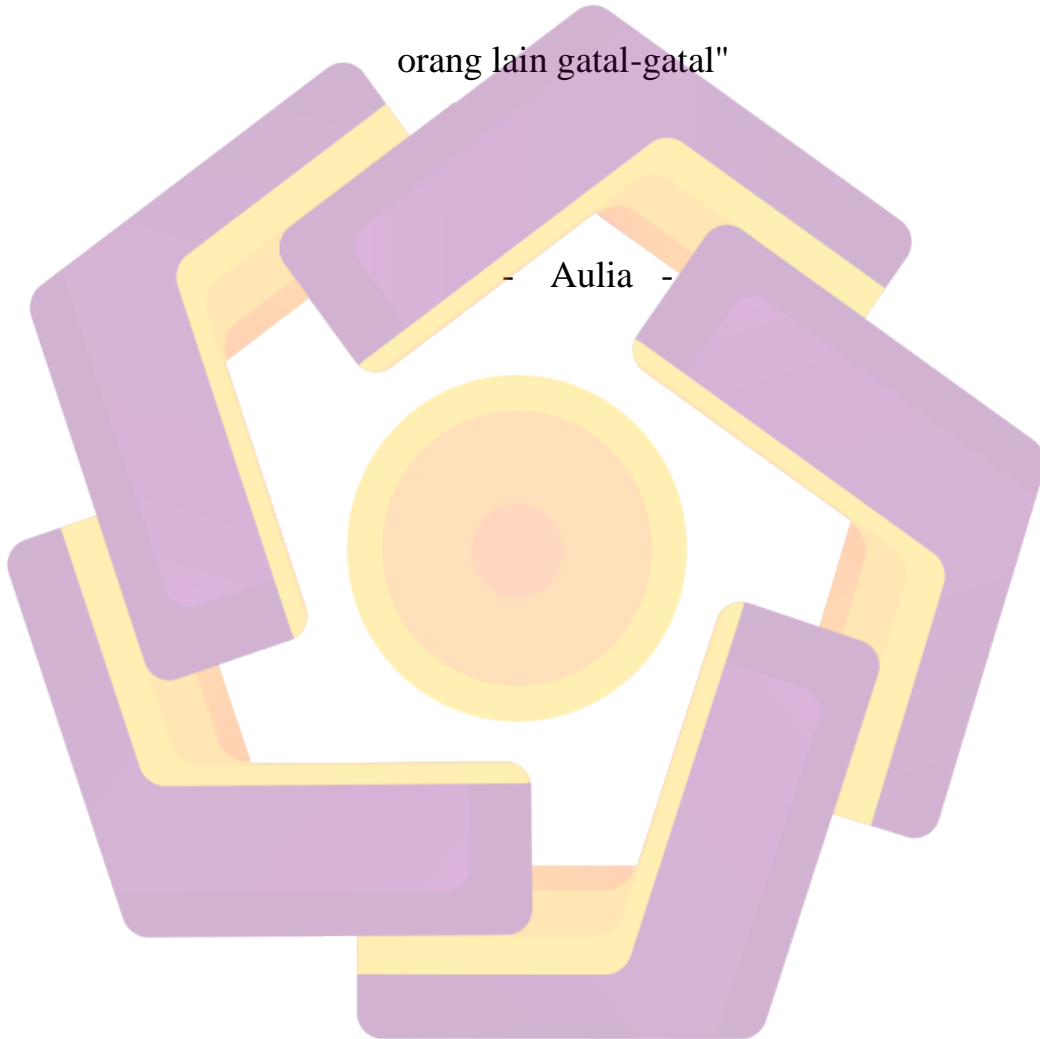
Yogyakarta, 30 Maret 2022



Aulia Rahma Damayanti
NIM. 17.11.1049

MOTTO

"Jadilah pribadi yang tekun, tapi jangan jadi pribadi yang ulet. Nanti
orang lain gatal-gatal"



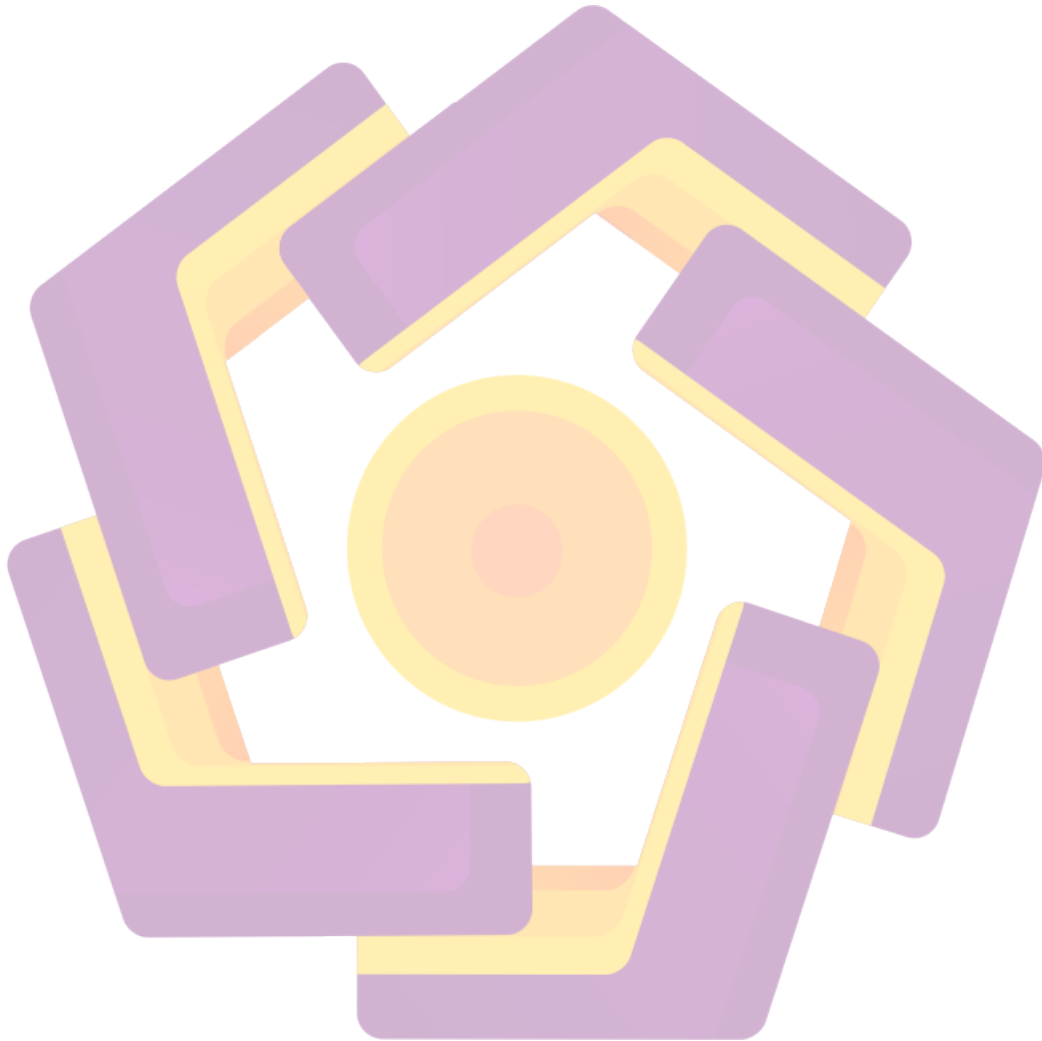
- Aulia -

PERSEMBAHAN

Puji syukur kita panjatkan kehadiran Allah SWT atas berkah dan karunia-Nya skripsi ini dapat terselesaikan dengan baik dan lancar. Dengan ini saya persembahkan skripsi ini untuk semua pihak yang terlibat langsung maupun tidak langsung, yaitu kepada:

1. Kepada orang tua (Pak Dwi dan Bu Is) yang terus memberikan semangat serta doa hingga saya dapat menyelesaikan tugas akhir/skripsi.
2. Dosen pembimbing sebelum ibu Mardhiya Hayati, M. Kom yang telah memberikan bimbingan kepada saya sebelumnya.
3. Dosen pembimbing bapak Agit Amrullah, S. Kom, M.Kom. yang terhormat senantiasa membimbing saya dari awal hingga akhir skripsi ini terselesaikan.
4. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan ilmu serta bimbingan kepada saya selama perkuliahan.
5. Keluarga Pak Temu Raharjo yang telah memberikan support dalam tempat tinggal dan apapun selama saya menempuh masa studi.
6. Teman-teman Universitas Amikom Yogyakarta yang khususnya kelas 17IF-02 yang telah menemani saya dan memberikan semangat serta membantu dalam pengerjaan skripsi ini.

7. Teman-teman Tuyak dan Sangkrah, Ana, Nia, Vita, Rere, Jody, Lupek, Ilham dan pacar saya Bagas Rifky yang selalu memberikan masukan dalam mengerjakan skripsi.



KATA PENGHANTAR

Puji syukur penulis ucapkan kepada Allah SWT, yang mana telah memberikan kesehatan dan karunia-Nya kepada penulis serta kekuatan untuk menyelesaikan skripsi yang berjudul **“IDENTIFIKASI TOPIC MODELLING PADA ULASAN PRODUK DENGAN ANALISIS LATENT SEMANTIC SIMILARITY”**. Tidak lupa penulis mengucapkan shalawat dan salam kepada junjungan Nabi Besar Muhammad SAW. Penyelesaian tulisan ini terlepas bantuan dari berbagai pihak yang terkait secara langsung maupun tidak langsung, terutama dan teristimewa dipersembahkan kepada kedua orang tua tercinta yang senantiasa memberikan rasa sayang, didikan, serta doa yang selalu di panjatkan pada Allah kepada penulis.

Skripsi ini dapat terselesaikan dengan bantuan berbagai pihak, maka dari itu penulis menyatakan rasa hormat dan terimakasih kepada:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom selaku ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Mardhiya Hayati, M. Kom. Selaku pembimbing sebelumnya yang telah memberikan masukan serta nasihat kepada saya.

5. Bapak Agit Amrullah, S. Kom., M. Kom. selaku pembimbing yang senantiasa memberikan masukan serta nasihat dalam penulisan skripsi ini.
6. Bapak Ali Mustopa, M.Kom dan Bapak Andy Sunyoto, M.Kom., Dr. selaku dosen penguji, terima kasih atas saran dan kritiknya sehingga penelitian ini menjadi lebih baik.

Penulis menyadari masih ada kekurangan dan kelemahan dalam pembuatan skripsi ini. Maka penulis mengharapkan adanya kritik dan saran dari segala pihak agar menambah kesempurnaan dalam skripsi ini.

Yogyakarta, 30 Maret 2022

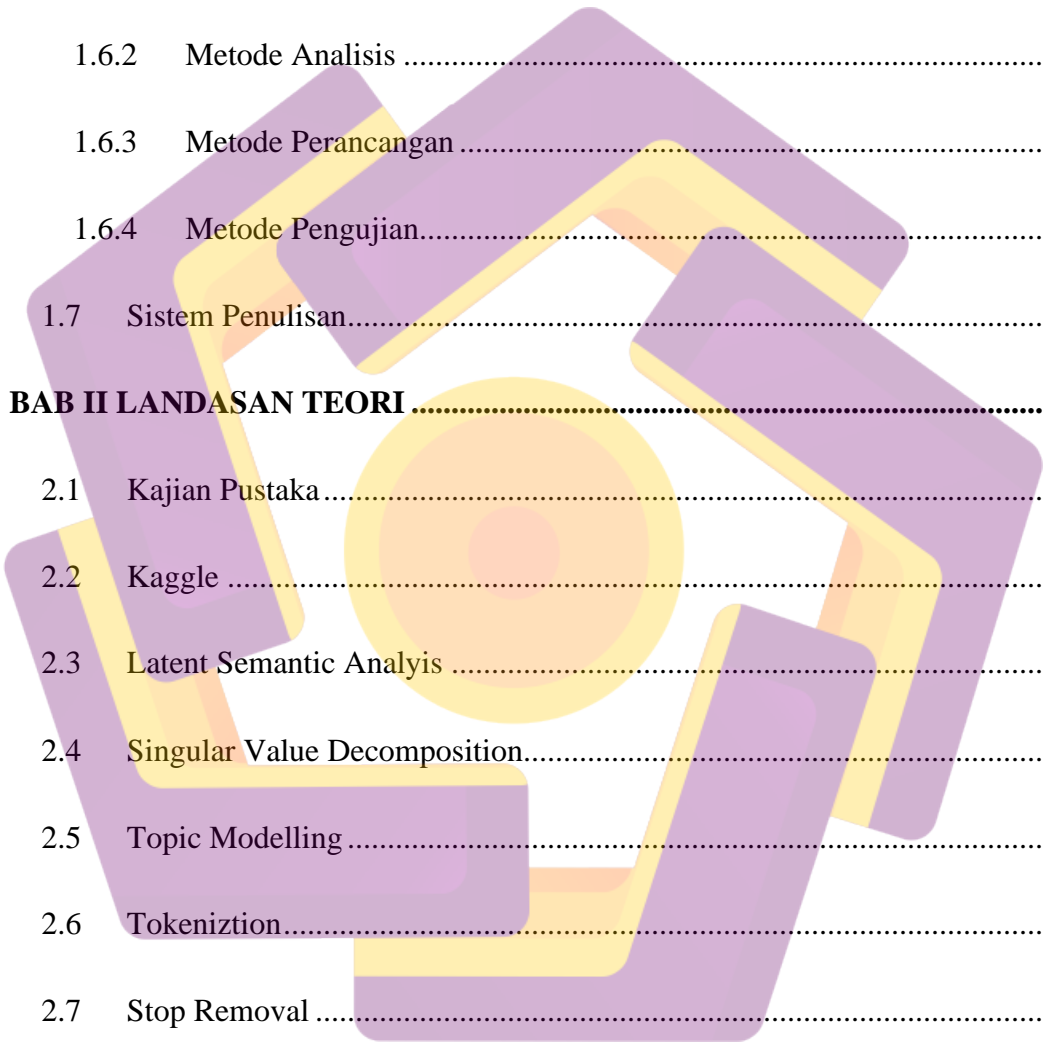


Aulia Rahma Damayanti

NIM 17.11.1049

DAFTAR ISI

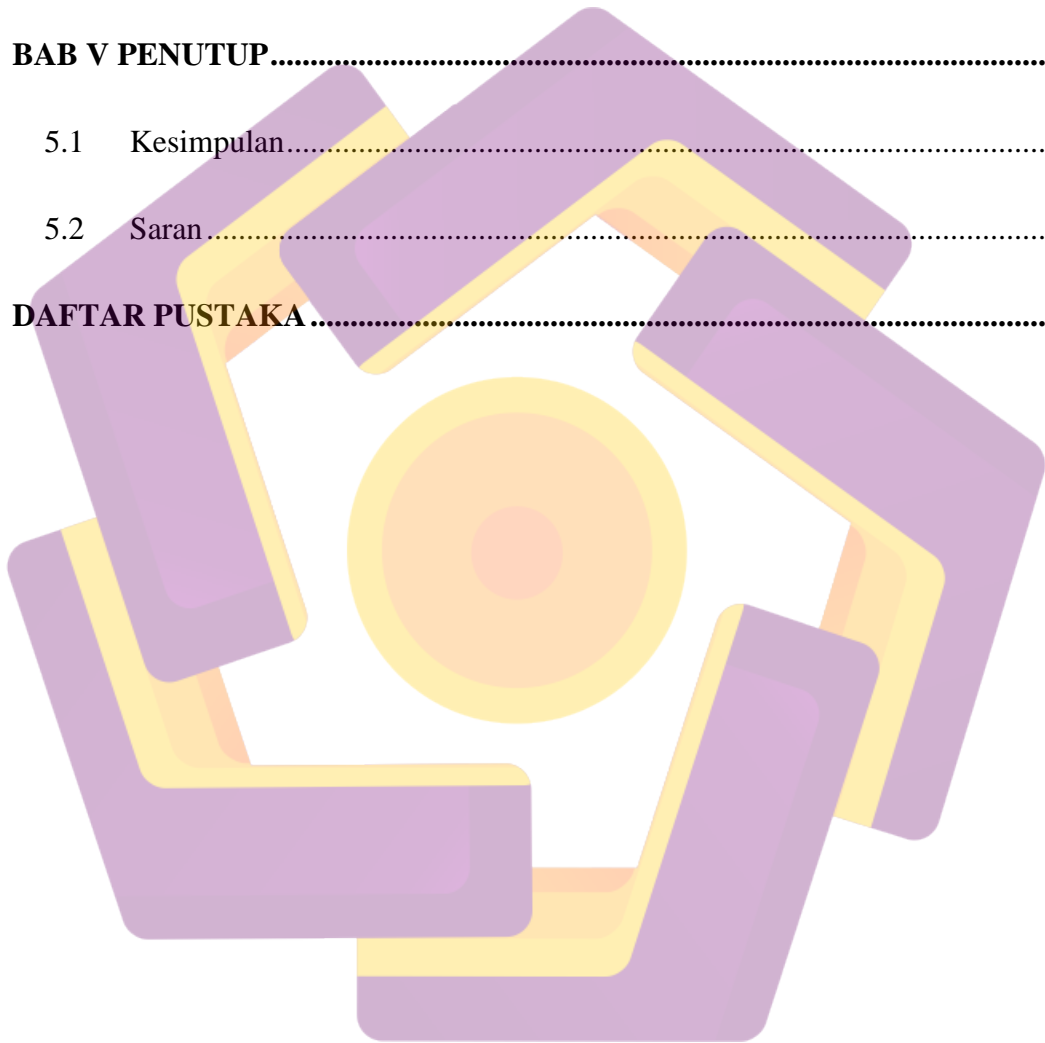
JUDUL	ii
PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGHANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvi
INTISARI	xvii
ABSTRACT	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian	4



1.5	Manfaat Penelitian.....	4
1.6	Metode Penelitian.....	4
1.6.1	Metode Pengumpulan Data.....	4
1.6.2	Metode Analisis	5
1.6.3	Metode Perancangan	5
1.6.4	Metode Pengujian.....	6
1.7	Sistem Penulisan.....	6
BAB II	LANDASAN TEORI	8
2.1	Kajian Pustaka.....	8
2.2	Kaggle	10
2.3	Latent Semantic Analysis	10
2.4	Singular Value Decomposition.....	12
2.5	Topic Modelling.....	13
2.6	Tokenization.....	14
2.7	Stop Removal	15
2.8	Stemming.....	15
2.9	Python.....	16
2.10	Natural Language Processing (NLP)	16

2.11	Text Mining	17
BAB III METODE PENELITIAN		18
3.1	Alat dan Bahan Penelitian	18
3.1.1	Perangkat Keras (Hardware)	18
3.1.2	Perangkat Lunak (Software)	18
3.2	Alur Penelitian.....	19
3.2.1	Data Topic Modelling	19
3.2.2	Data Cleaning.....	20
3.2.3	Preprocessing.....	20
3.2.4	Feature Selection.....	22
3.2.5	Term-Document Matrix	22
3.2.6	Evaluasi.....	23
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Pengumpulan Data	24
4.2	Preprocessing.....	24
4.3	Processing.....	41
4.3.1	Document Term-Matrix	41
4.3.2	Topic Modelling LSA	42

4.3.3	Singular Value Decomposition	44
4.3.4	Visualisasi Topik Modelling LSA	45
4.3	Evaluasi	46
BAB V PENUTUP	48
5.1	Kesimpulan	48
5.2	Saran	48
DAFTAR PUSTAKA	xlix

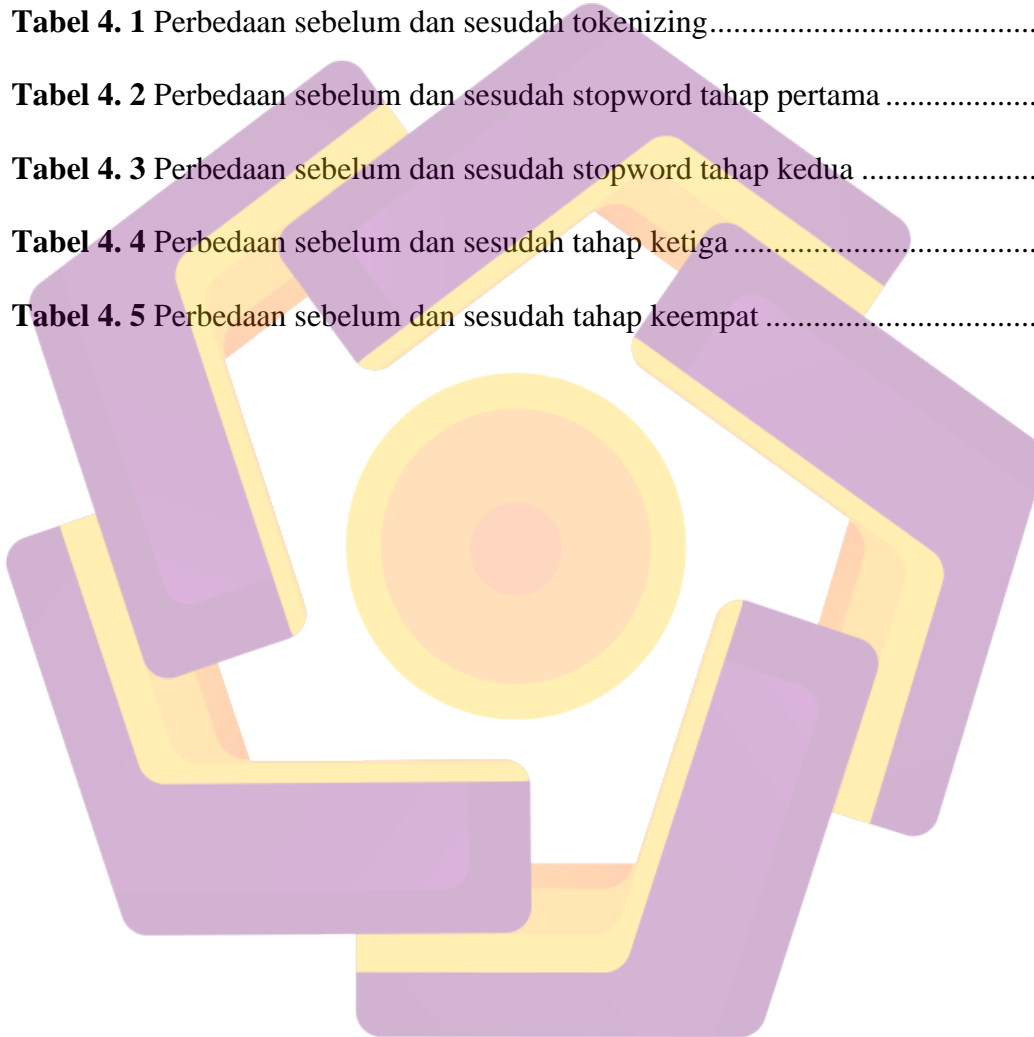


DAFTAR GAMBAR

Gambar 2. 1 Topic Modelling using LSA.....	11
Gambar 2. 2 Proses Singular Value Decomposition (SVD).....	12
Gambar 3. 1 Diagram Alur Tahapan Penelitian.....	19
Gambar 4. 1 Dataset.....	24
Gambar 4. 2 Source code document term matrix.....	41
Gambar 4. 3 Hasil vectorizer tf-idf.....	42
Gambar 4. 4 Hasil topik dengan metode LSA.....	43
Gambar 4. 5 Kumpulan dokumen topik.....	43
Gambar 4. 6 Hasil proses SVD.....	44
Gambar 4. 7 Visualisasi dokumen topik.....	45

DAFTAR TABEL

Tabel 2. 1 Contoh Tokenization	14
Tabel 2. 2 Contoh Proses Stopword Removal.....	15
Tabel 4. 1 Perbedaan sebelum dan sesudah tokenizing.....	26
Tabel 4. 2 Perbedaan sebelum dan sesudah stopwords tahap pertama	31
Tabel 4. 3 Perbedaan sebelum dan sesudah stopwords tahap kedua	33
Tabel 4. 4 Perbedaan sebelum dan sesudah tahap ketiga	35
Tabel 4. 5 Perbedaan sebelum dan sesudah tahap keempat	38



INTISARI

Platform digital saat ini sudah mengubah cara berinteraksi juga mengubah cara dalam menyampaikan pendapat atau ulasan. Sebagai contoh yaitu media e-commerce dengan menuliskan ulasan konsumen terhadap produk yang dibelinya. setiap ulasan pengguna memiliki peran penting dalam mempengaruhi tingkat penjualan dan minat konsumen terhadap produk yang dibelinya. Oleh karena itu, analisis semantic similarity dapat mencari ulasan mana saja yang memiliki maksud dan tujuan yang sama.

Penelitian ini menggunakan metode Latent Semantic Analysis (LSA). LSA adalah Teknik dalam pemrosesan Bahasa alami, khususnya dalam vector semantik. Terdapat beberapa metode yang akan dilakukan dalam penelitian ini, tahap pertama metode pengumpulan data dengan mengambil dataset dari Kaggle, metode analisis melakukan analisis mengenai metode LSA dalam mengidentifikasi pemodelan topik, metode perancangan mencakup proses preprocessing seperti data cleaning, tokenization, stopword removal dan stemming, metode pengujian tahap terakhir dalam penelitian untuk menghitung representasi nilai dari metode LSA pada pemodelan topik.

Metode Topic modelling dengan Latent semantic analysis (LSA) dapat menghasilkan data topik pada data ulasan produk e-commerce yang diambil dari kaggle. Dengan menghasilkan 10 topik teratas dari pemrosesan. Dalam bentuk visualisasi ditampilkan bahwa data 10 topik yang muncul terdapat 2 topik yang nilai topic modellingnya tinggi yaitu Topik 0 dengan nilai 19,89 dan Topik 4 dengan nilai 19,75.

Kata Kunci : *Latent Semantic Analysis (LSA), SVD, Ulasan Produk.*

ABSTRACT

Today's digital platforms are changing the way we interact and also changing the way we express opinions or reviews. For example, e-commerce media with consumer reviews of the products they buy. Each user review has an important role in influencing the level of sales and consumer interest in the products they buy. Therefore, semantic similarity analysis can look for any reviews that have the same goals and objectives.

*This study uses the **Latent Semantic Analysis (LSA)** method. LSA is a technique for expressing natural language, especially in **vector semantics**. There are several methods that will be carried out in this study, the first stage of data collection by taking datasets from Kaggle, the analysis method regarding the LSA method in identifying modeling topics, the design method includes preprocessing processes such as data cleaning, tokenization, stopword removal and stemming, the last stage testing method. in research to calculate the value representation of the LSA method on topic modeling.*

*Topic modeling method with **Latent semantic analysis (LSA)** can generate topic data on e-commerce product data reviews taken from kaggle. By generating top 10 topics from ideas. In the form of visualization, it is shown that in the data of 10 topics that appear, there are 2 topics that have a high value of modeling topics, namely Topic 0 with a value of 19.89 and Topic 4 with a value of 19.75.*

Keyword : Latent Semantic Analysis (LSA), SVD, ReviewProduct.