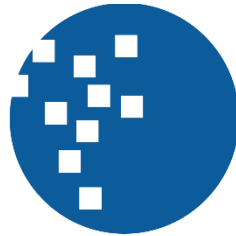


**ANALISIS SENTIMEN TERHADAP REVIEW QUORA
INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN
DECISION TREE DAN K-NEAREST NEIGHBORS**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

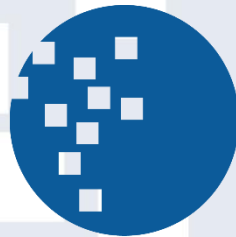
Marcellius Cahyadi

00000042599

**PROGRAM STUDI SISTEM INFORMASI
FALKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2024

**ANALISIS SENTIMEN TERHADAP REVIEW QUORA
INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN
DECISION TREE DAN K-NEAREST NEIGHBORS**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Marcellius Cahyadi

00000042599

PROGRAM STUDI SISTEM INFORMASI

FALKULTAS TEKNIK DAN INFORAMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Marcellius Cahyadi

Nomor Induk Mahasiswa : 00000042599

Program studi : Sistem informasi

Skripsi dengan judul:

ANALISIS SENTIMEN TERHADAP REVIEW QUORA INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN DECISION TREE DAN K-NEAREST NEIGHBORS

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 14 Mei 2024



Marcellius Cahyadi

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Skripsi dengan judul

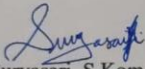
ANALISIS SENTIMEN TERHADAP REVIEW QUORA INDONESIA DI
GOOGLE PLAY STORE MENGGUNAKAN DECISION TREE DAN K-
NEAREST NEIGHBORS

Oleh

Nama : Marcellius Cahyadi
NIM : 00000042599
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Senin, 3 Juni 2024
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang


Suryasari, S.Kom., M.T. 13/6-2024.
0323088301

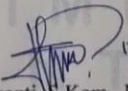
Penguji


Johan Setiawan, S.Kom., MM. 13/6/24
0327106402

Pembimbing


Ririn Ikana Desanti, S.Kom., M.Kom. 13/6/24
NIDN: 0313058001

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom. 13/6/24

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Marcellius Cahyadi
NIM : 00000042599
Program Studi : Sistem Informasi
Jenjang : S1
Judul Karya Ilmiah :

ANALISIS SENTIMEN TERHADAP REVIEW QUORA INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN DECISION TREE DAN K-NEAREST NEIGHBORS

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial. Saya tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: Dalam proses pengajuan penerbitan ke dalam jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 7 Juni 2024



Marcellius Cahyadi

* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji Syukur atas selesainya penulisan skripsi ini dengan judul: “ANALISIS SENTIMEN TERHADAP REVIEW QUORA INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN DECISION TREE DAN K-NEAREST NEIGHBORS” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Strata 1 Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Diharapkan skripsi ini dapat menjadi ilmu yang bermanfaat bagi masyarakat. Oleh karena itu, saya mengucapkan terima kasih kepada:

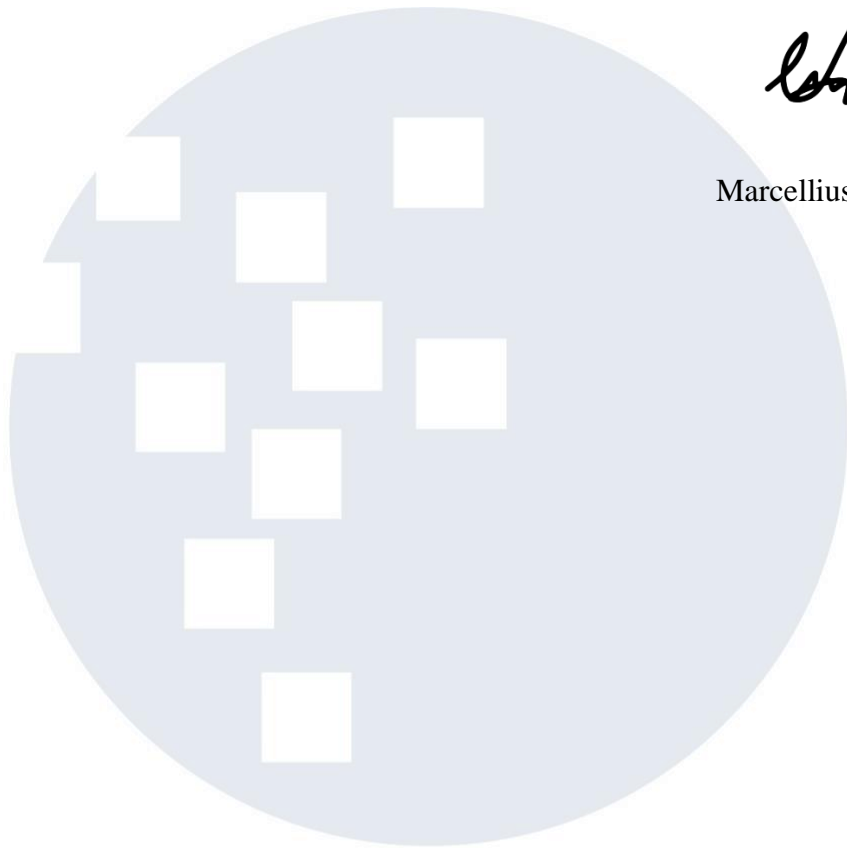
1. Bapak Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat menjadi sumber informasi demi perkembangan ilmu pengetahuan yang bermanfaat bagi pembaca dan masyarakat.

Tangerang, 13 Mei 2024



Marcellius Cahyadi



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ANALISIS SENTIMEN TERHADAP REVIEW QUORA INDONESIA DI GOOGLE PLAY STORE MENGGUNAKAN DECISION TREE DAN K-NEAREST NEIGHBORS

Marcellius Cahyadi

ABSTRAK

Quora merupakan aplikasi sosial media yang berbentuk tanya jawab. Pengguna dapat mengajukan pertanyaan yang dapat dijawab oleh pengguna Quora lain dengan jawaban yang beragam tergantung dengan latar belakang pengguna tersebut. Aplikasi Quora memiliki banyak pengguna sekitar 50 juta pengguna dan telah direview oleh 970 ribu pengguna dengan rata-rata rating sebesar 4.7. Hal tersebut membuat aplikasi ini cukup banyak digunakan namun aplikasi ini tidak terlalu terdengar. Untuk itu, penelitian ini memiliki tujuan untuk melihat minat dan respon masyarakat Indonesia terhadap aplikasi Quora melalui analisis sentimen menggunakan *Jupyter Notebook* dengan algoritma *K-Nearest Neighbors* dan *Decision tree*. Proses analisis sentimen ini terdiri dari beberapa alur yaitu pemilihan topik *data*, *data scrapping*, *data labeling*, *data processing*, *TF-IDF*, *Confusion Matrix*, *data Mining*, dan *Deployment* yang berupa visualisasi. Hasil penelitian mendapati bahwa pengguna aplikasi Quora di Indonesia mendapat respon yang positif ditandai dengan jumlah 1234 atau sekitar 82,3% pengguna sementara untuk respons negatif hanya berkisar 266 ulasan atau hanya berkisar 17,7%. Algoritma terbaik yang dapat digunakan adalah *K-Nearest Neighbors* dengan pembagian 70:30. Hasil didapatkan oleh algoritma yaitu nilai akurasi sebesar 83%, *Precision* sebesar 84%, *Recall* sebesar 97%, *F1_Score* sebesar 90%, dan *Error_Rate* sebesar 16%.

Kata kunci: Analisis Sentimen, Minat, Respon, Quora, Visualisasi

Sentiment Analysis of Quora Indonesia Reviews on Google Play Store using Decision Tree dan K-Nearest Neighbors

Marcellius Cahyadi

ABSTRACT (English)

Quora is a social media application in the form of questions and answers. Users can ask questions that can be answered by other Quora users with varying answers depending on the user's background. The Quora application has around 50 million users and has been reviewed by 970 thousand users with an average rating of 4.7. This makes this application quite widely used, but this application is not very popular. For this reason, this research aims to see the interest and response of Indonesian people to the Quora application through sentiment analysis using Jupyter Notebook with K-Nearest Neighbors and Decision trees algorithms. The sentiment analysis process consists of several flows, namely selecting data topics, data scrapping, data labeling, data processing, TF-IDF, Confusion Matrix, data Mining, and Deployment in the form of visualization. The research results found that users of the Quora application in Indonesia received positive responses marked by 1234 or around 82.3% of users, while negative responses were only around 266 reviews or only around 17.7%. The best algorithm that can be used is K-Nearest Neighbors with a 70:30 division. The results obtained by the algorithm are an accuracy value of 83%, Precision of 84%, Recall of 97%, F1_Score of 90%, and Error_Rate of 16%.

Keywords: *Interest, Quora, Sentiment Analysis, Responses, Visualization*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA.....	iv
KATA PENGANTAR	v
ABSTRAK.....	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan dan Manfaat Penelitian.....	5
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan.....	6
BAB II LANDASAN TEORI.....	9
2.1 Penelitian Terdahulu.....	9
2.2 Teori yang digunakan untuk Pengujian.....	16
2.2.1 Analisis Sentimen.....	16
2.2.2 Teori Quora.....	16

2.2.3	<i>Sentiment Labelling</i>	16
2.2.4	<i>Text Mining</i>	17
2.3	Algoritma yang Digunakan untuk Pengujian	18
2.3.1	Teori Decision Tree.....	18
2.3.2	Teori K-Nearest Neighbors	18
2.3.3	TF-IDF (<i>Term Frequency - Inverse Document Frequency</i>)	18
2.4	<i>Tools</i> yang digunakan untuk Pengujian.....	19
2.4.1	<i>Jupyter Notebook</i>	19
2.4.2	Python	20
BAB III METODOLOGI PENELITIAN.....		21
3.1	Gambaran Umum Objek Penelitian.....	21
3.2	Metode Penelitian.....	21
3.3	Teknik Pengumpulan Data	22
3.4	Teknik Pengambilan Sampel.....	22
3.5	Alur Penelitian.....	22
3.5.1	Pemilihan Topik <i>Data</i>	24
3.5.2	<i>Data Scrapping</i>	24
3.5.3	<i>Data Labeling</i>	25
3.5.4	<i>Data Preprocessing</i>	25
3.5.5	TF-IDF	26
3.5.6	<i>Confusion Matrix</i>	26
3.5.7	<i>Data Mining</i>	27
3.5.8	<i>Deployment</i>	27
3.6	Teknik <i>Data Mining</i>	28
3.6.1	<i>Business Understanding</i>	29

3.6.2	<i>Data Understanding</i>	29
3.6.3	<i>Data Preparation</i>	29
3.6.4	<i>Modelling</i>	29
3.6.5	<i>Evaluation</i>	30
3.6.6	<i>Deployment</i>	30
BAB IV ANALISIS DAN HASIL PENELITIAN		31
4.1	Business Understanding	31
4.1.1	<i>Software</i>	33
4.2	<i>Data Understanding</i>	34
4.2.1	<i>Data Scrapping</i>	34
4.3	<i>Data Preparation</i>	39
4.3.1	<i>Data Processing</i>	39
4.4	<i>Data Modelling</i>	50
4.4.1	TF-IDF	50
4.4.2	<i>Data Splitting</i>	51
4.4.3	Hasil Pengujian	54
4.5	<i>Evaluation</i>	62
4.6	<i>Deployment</i>	63
BAB V KESIMPULAN DAN SARAN		69
5.1	Kesimpulan	69
5.2	Saran	70
5.2.1	Limitasi Masalah yang Diselesaikan	70
DAFTAR PUSTAKA		71
LAMPIRAN		75

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	9
Tabel 4. 1 Contoh dari Penghilangan Mention	43
Tabel 4. 2 Contoh Penghilangan Link.....	43
Tabel 4. 3 Contoh dari penghilangan tanda baca	44
Tabel 4. 4 Contoh dari Penggunaan Case Folding	44
Tabel 4. 5 Contoh dari Penghilangan Spasi Berlebih	45
Tabel 4. 6 Contoh dari Penghilangan Hashtag	46
Tabel 4. 7 Hasil dari Stopwords.....	48
Tabel 4. 8 Hasil dari Tokenizing.....	48
Tabel 4. 9 Hasil dari Stemming.....	49
Tabel 4. 10 Tabel Confusion Matrix K-Nearest Neighbor dengan 90:10.....	54
Tabel 4. 11 Tabel Hasil dari K-Nearest Neighbors 90:10.....	55
Tabel 4. 12 Tabel Confusion Matrix K-Nearest Neighbor dengan 80:20.....	56
Tabel 4. 13 Tabel Hasil dari K-Nearest Neighbors 80:20.....	56
Tabel 4. 14 Tabel Confusion Matrix K-Nearest Neighbor dengan 70:30.....	57
Tabel 4. 15 Tabel Hasil dari K-Nearest Neighbors 70:30.....	58
Tabel 4. 16 Tabel Confusion Matrix Decision Tree dengan 90:10.....	58
Tabel 4. 17 Tabel Hasil dari Decision Tree 90:10	59
Tabel 4. 18 Tabel Confusion Matrix Decision Tree dengan 80:20.....	60
Tabel 4. 19 Tabel Hasil dari Decision Tree 80:20	60
Tabel 4. 20 Tabel Confusion Matrix Decision Tree dengan 70:30.....	61
Tabel 4. 21 Tabel Hasil dari Decision Tree 70:30	61
Tabel 4. 22 Total Hasil dari K-Nearest Neighbor	62
Tabel 4. 23 Tabel Hasil dari Decision Tree	62
Tabel 4. 24 Perbandingan Algoritma K-Nearest Neighbors dan Decision Tree... 63	63

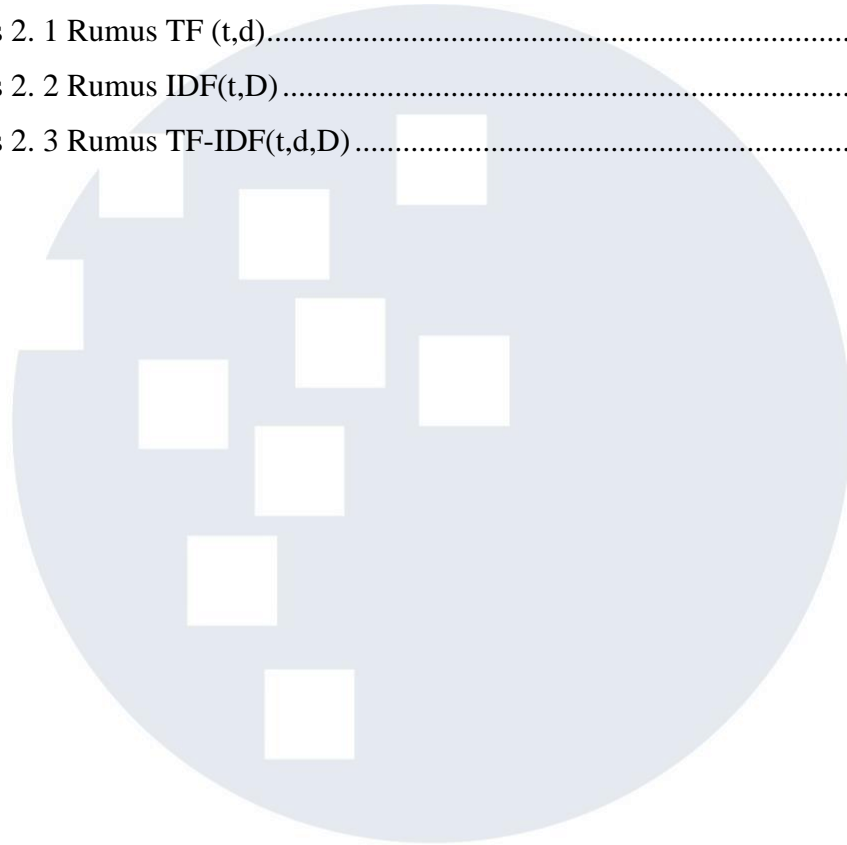
DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian.....	23
Gambar 3. 2 Diagram Alur CISDM.....	28
Gambar 4. 1 Review Quora di Google Play Store	32
Gambar 4. 2 Alur Data Understanding	34
Gambar 4. 3 Library Google Play Scrapper.....	35
Gambar 4. 4 Kode untuk scrapping data.....	35
Gambar 4. 5 Tampilan data review mentah	36
Gambar 4. 6 Jumlah data yang telah discrapping	36
Gambar 4. 7 Hasil sortir data	37
Gambar 4. 8 Jumlah Null pada Data	38
Gambar 4. 9 Kode untuk Import csv.....	38
Gambar 4. 10 Isi dari File csv yang Telah discrapping.....	39
Gambar 4. 11 Codingan Labeling sentiment.....	40
Gambar 4. 12 Hasil table setelah labeling sentimen	40
Gambar 4. 13 Total hasil sentimen positif dan negatif	41
Gambar 4. 14 Library untuk pemrosesan data	42
Gambar 4. 15 Kode untuk menghilangkan mention	43
Gambar 4. 16 Kode untuk menghilangkan link	43
Gambar 4. 17 Kode dari penghilangan tanda baca	44
Gambar 4. 18 Kode dari Case Folding.....	45
Gambar 4. 19 Kode dari Penghilangan Spasi Berlebih.....	45
Gambar 4. 20 Kode dari Penghilangan Hashtag	46
Gambar 4. 21 Kode dari Tokenizing.....	46
Gambar 4. 22 Hasil dari Tokenizing	46
Gambar 4. 23 Kode untuk Stopwords	47
Gambar 4. 24 Hasil kode untuk menghilangkan Stopwords.....	47
Gambar 4. 25 Kode untuk Stemming.....	49
Gambar 4. 26 Hasil dari Stemming.....	49
Gambar 4. 27 Kode dari TD-IDF.....	51
Gambar 4. 28 Hasil dari TD-IDF	51

Gambar 4. 29 Kode untuk Skenario Pembagian 90:10.....	52
Gambar 4. 30 Jumlah Data Training dan Data Testing 90:10 berdasarkan Sentimen	52
Gambar 4. 31 Jumlah Data untuk Skenario Pembagian 80:20.....	52
Gambar 4. 32 Jumlah Data Training dan Data Testing 80:20 berdasarkan Sentimen	53
Gambar 4. 33 Jumlah Data untuk Skenario Pembagian 70:30.....	53
Gambar 4. 34 Jumlah Data Training dan Data Testing 70:30 berdasarkan Sentimen	54
Gambar 4. 35 Kode untuk Confusion Matrix K-Nearest Neighbor dengan 90:10	54
Gambar 4. 36 Kode dan Hasil dari K-Nearest Neighbors dengan 90:10.....	55
Gambar 4. 37 Kode untuk Confusion Matrix K-Nearest Neighbor dengan 80:20	55
Gambar 4. 38 Kode dan Hasil dari K-Nearest Neighbors dengan 80:20.....	56
Gambar 4. 39 Kode untuk Confusion Matrix K-Nearest Neighbor dengan 70:30	57
Gambar 4. 40 Kode dan Hasil dari K-Nearest Neighbors dengan 70:30.....	57
Gambar 4. 41 Kode untuk Confusion Matrix Decision Tree dengan 90:10	58
Gambar 4. 42 Kode dan Hasil dari Decision Tree dengan 90:10.....	59
Gambar 4. 43 Kode untuk Confusion Matrix Decision Tree dengan 80:20	59
Gambar 4. 44 Kode dan Hasil dari Decision Tree dengan 80:20.....	60
Gambar 4. 45 Kode untuk Confusion Matrix Decision Tree dengan 70:30	61
Gambar 4. 46 Kode dan Hasil dari Decision Tree dengan 70:30.....	61
Gambar 4. 47 Barplot Jumlah Review berdasarkan Bintang	63
Gambar 4. 48 Barplot Sentimen.....	64
Gambar 4. 49 Penyebaran Persentase Sentimen	65
Gambar 4. 50 WordCloud Sentimen Positif	65
Gambar 4. 51 Wordcloud Sentimen Negatif.....	66
Gambar 4. 52 Heatmap Confusion Matrix untuk K-Nearest Neighbors.....	66
Gambar 4. 53 Heatmap Untuk Decision Tree.....	67
Gambar 4. 54 Hasil Deployment Website Sentimen Positif	68
Gambar 4. 55 Hasil Deployment Website Sentimen Negatif	68

DAFTAR RUMUS

Rumus 2. 1 Rumus TF (t,d).....	19
Rumus 2. 2 Rumus IDF(t,D).....	19
Rumus 2. 3 Rumus TF-IDF(t,d,D).....	19

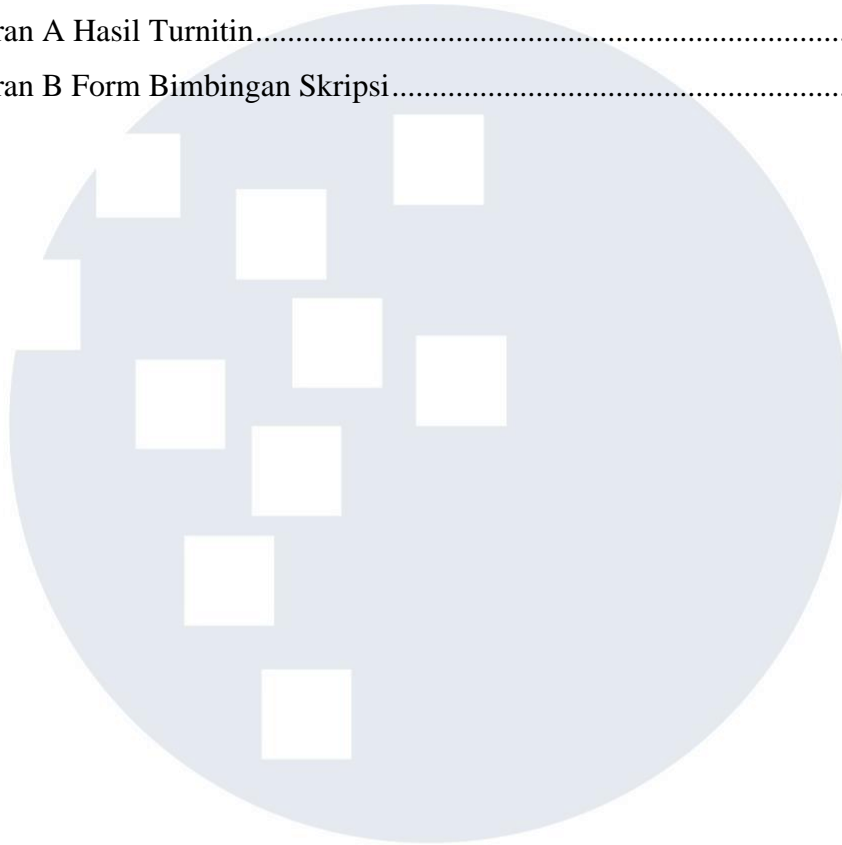


UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Hasil Turnitin.....	75
Lampiran B Form Bimbingan Skripsi.....	86



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA