

TUGAS AKHIR

OPTIMASI K-MEANS MENGGUNAKAN
ALGORITMA GENETIKA UNTUK
PENGELOMPOKAN POPULARITAS WEBTOON



NABILA RASYA PUTRI AMANI

20110023

PROGRAM STUDI S1 SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO

2024

TUGAS AKHIR

OPTIMASI K-MEANS MENGGUNAKAN ALGORITMA GENETIKA UNTUK PENGELOMPOKAN POPULARITAS WEBTOON

K-MEANS OPTIMIZATION USING GENETIC ALGORITHM FOR CLUSTERING WEBTOONS

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Sains Data



NABILA RASYA PUTRI AMANI

20110023

**PROGRAM STUDI S1 SAINS DATA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2024**

HALAMAN PERSETUJUAN PEMBIMBING

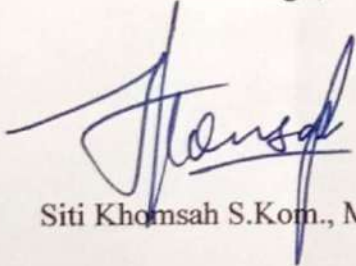
**OPTIMASI K-MEANS MENGGUNAKAN ALGORITMA
GENETIKA UNTUK PENGELOMPOKAN POPULARITAS
WEBTOON**

***K-MEANS OPTIMIZATION USING GENETIC ALGORITHM
FOR CLUSTERING WEBTOON***

Dipersiapkan dan Disusun oleh
NABILA RASYA PUTRI AMANI
20110023

**Fakultas Informatika
Institut Teknologi Telkom Purwokerto
Pada Tanggal 4 Juni 2024**

Pembimbing I,



Siti Khomsah S.Kom., M.Cs.

NIDN. 0517108101

Pembimbing II,



Aina Latifa Riyana Putri, S.Si., M.Mat.

NIDN. 0629019901

HALAMAN PENGESAHAN

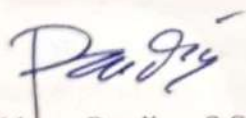
OPTIMASI K-MEANS MENGGUNAKAN ALGORITMA GENETIKA UNTUK PENGELOMPOKAN POPULARITAS WEBTOON

K-MEANS OPTIMIZATION USING GENETIC ALGORITHM FOR CLUSTERING WEBTOONS

Disusun oleh
NABILA RASYA PUTRI AMANI
20110023

Telah Diujikan Dan Dipertahankan Dalam Sidang Ujian Sidang Akhir Pada
Hari Rabu, 25 Juni 2024

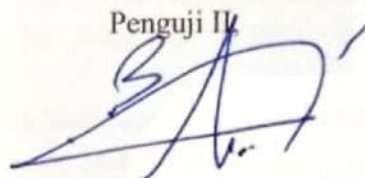
Penguji I,



Dr. Ridwan Pandiya, S.Si., M.Sc.

NIDN. 0625088202

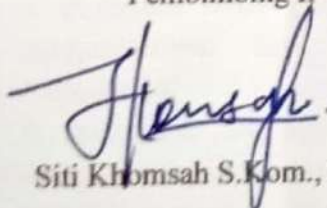
Penguji II



Andi Prademon Yunus, ST., M.Sc.Eng., Ph.D.

NIDN. 0616129301

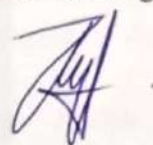
Pembimbing I,



Siti Khomsah S.Kom., M.Cs.

NIDN. 0517108101

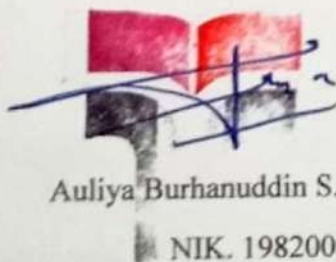
Pembimbing II,



Aina Latifa Riyana Putri, S.Si., M.Mat.

NIDN : 0629019901

Dekan,



Auliya Burhanuddin S.Si., M.Kom
NIK. 19820008

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nabila Rasya Putri Amani
NIM : 20110023
Program Studi : S1 Sains Data

Menyatakan bahwa Tugas Akhir dengan judul berikut:

OPTIMASI K-MEANS MENGGUNAKAN ALGORITMA GENETIKA UNTUK PENGELOMPOKAN POPULARITAS WEBTOON

Dosen Pembimbing Utama : Siti Khomsah S.Kom., M.Cs

Dosen Pembimbing Pendamping : Aina Latifa Riyana Putri, S.Si., Mat

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim 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 tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian 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.

Purwokerto, 4 Juni 2024

Yang Menyatakan,



Nabila Rasya Putri Amani

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan tugas akhir dengan judul **“OPTIMASI K-MEANS MENGGUNAKAN ALGORITMA GENETIKA UNTUK PENGELOMPOKAN POPULARITAS WEBTOON”**. Dalam penyusunan tugas akhir ini tidak terlepas dari bantuan serta dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang yang senantiasa memberikan petunjuk dan kekuatan kepada peneliti dalam menyelesaikan tugas akhir ini.
2. Kedua orang tua peneliti, Winarko dan Siti Marjati yang telah memberi kasih sayang, doa, nasehat, serta usaha yang tidak akan mampu peneliti balas hingga kapanpun. Semoga penulis mampu menjadi anak yang dapat kalian banggakan
3. Ibu Dr. Tenia Wahyuningrum, S.Kom., M.T. selaku Rektor Institut Teknologi Telkom Purwokerto.
4. Bapak Auliya Burhanuddin S.Si., M.Kom. selaku Dekan Fakultas Informatika Institut Teknologi Telkom Purwokerto, sekaligus dosen penguji 1 yang telah meluangkan waktu ditengah kesibukan beliau untuk menguji kelayakan Tugas Akhir yang telah diselesaikan oleh peneliti.
5. Ibu Siti Khomsah, S.Kom., M.Cs selaku Kaprodi Sains Data, Dosen Wali sekaligus Dosen Pembimbing Utama yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, masukan dan arahan kepada peneliti dalam menempuh studi di Prodi Sains Data hingga proses penyelesaian Tugas Akhir ini.
6. Ibu Aina Latifa Riyana Putri, S.Si., Mat selaku dosen pembimbing pendamping yang telah meluangkan waktu ditengah kesibukan beliau, memberikan kritik, masukan dan arahan kepada peneliti dalam proses penyelesaian Tugas Akhir ini.

7. Seluruh jajaran staf dosen di Fakultas Informatika yang telah memberi ilmu pengetahuan yang sangat berharga selama peneliti menempuh pendidikan di Fakultas Informatika Institut Teknologi Telkom Purwokerto.
8. Adik peneliti, Abiyyu Duta Wimar An Nafi'. Semoga penulis dapat menjadi contoh seorang kakak yang baik untuk kalian.
9. Seluruh teman seperjuangan yang tidak mampu peneliti sebutkan seluruhnya pada bagian kata pengantar ini. Untuk semua pengalaman bermanfaat yang peneliti rasakan selama menempuh pendidikan di Fakultas Informatika Institut Teknologi Telkom Purwokerto.

Purwokerto, 4 Juni 2024



Nabila Rasya Putri Amani

DAFTAR ISI

TUGAS AKHIR.....	ii
HALAMAN PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
ABSTRAK	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	16
1.1 Latar Belakang	16
1.2 Rumusan Masalah	19
1.3 Pertanyaan Penelitian	19
1.4 Batasan Masalah.....	19
1.5 Tujuan Penelitian.....	20
1.6 Manfaat Penelitian.....	20
BAB II TINJAUAN PUSTAKA	21
2.1 Kajian Pustaka	21
2.2 Dasar Teori	29
2.2.1 LINE Webtoon	29
2.2.2 Data Mining	29
2.2.3 <i>Clustering</i>	30
2.2.4 K-Means.....	31
2.2.5 Algoritma Genetika.....	33

2.2.6	<i>Silhouette Score</i>	35
2.2.7	Davies Bouldin Index	36
BAB III METODOLOGI PENELITIAN.....		38
3.1	Subjek dan Objek Penelitian	38
3.2	Alat dan Bahan Penelitian	38
3.2.1	Alat Penelitian.....	38
3.2.2	Bahan.....	38
3.3	Diagram Alir Penelitian.....	39
BAB IV HASIL DAN PEMBAHASAN.....		43
4.1	Pre-processing	43
4.1.1	Seleksi Fitur	45
4.1.2	Mengubah variabel/kolom	46
4.1.3	Deskripsi Data.....	47
4.1.4	Identifikasi <i>Outliers</i>	50
4.1.5	Menghapus <i>outlier</i>	52
4.2	Modeling	53
4.2.1	K-Means	53
4.2.2	Kombinasi K-Means dan <i>Genetic Algorithm</i>	58
4.3	Perbandingan Klustering	63
4.3.1	Hasil Klustering dengan $K = 2$ menggunakan K-Means dan Kombinasi K-Means + <i>Genetic Algorithm</i>	63
4.3.2	Hasil Klustering dengan $K = 3$ menggunakan K-Means dan Kombinasi K-Means + <i>Genetic Algorithm</i>	63
4.4	Hasil Evaluasi Klustering.....	64
4.4.1	Data Webtoon.....	64
4.4.2	Data Dummy.....	65
BAB V KESIMPULAN DAN SARAN.....		67
5.1	Kesimpulan.....	67
5.2	Saran	68
DAFTAR PUSTAKA		69
LAMPIRAN.....		74

DAFTAR TABEL

Tabel 2. 1 Kajian Pustaka.....	24
Tabel 4. 1 Deskripsi Informasi Tabel	43
Tabel 4. 2 Dataset Komik Webtoon.....	44
Tabel 4. 3 Tipe Data pada Dataset.....	45
Tabel 4. 4 Dataset Setelah Pemilihan Kolom.....	45
Tabel 4. 5 Tipe Data pada Dataset.....	46
Tabel 4. 6 Deskripsi Data	47
Tabel 4. 7 Data Dummy	48
Tabel 4. 8 Perbandingan Outlier.....	52
Tabel 4. 9 Titik Centroid K-Means K = 2	53
Tabel 4. 10 Titik Centroid K-Means K = 3	56
Tabel 4. 11 Parameter Genetic Algorithm K = 2.....	59
Tabel 4. 12 Centroid K-Means + Genetic Algorithm K = 2.....	59
Tabel 4. 13 Parameter Genetic Algorithm K = 3.....	61
Tabel 4. 14 Centroid K-Means + Genetic Algorithm K = 3.....	61
Tabel 4. 15 Hasil Klaster K-Means & Genetic Algorithm.....	64
Tabel 4. 16 Hasil Klaster K-Means & Genetic Algorithm Data Dummy	65
Tabel 4. 17 Hasil Klaster K-Means & Genetic Algorithm Data Dummy	65
Tabel 4. 18 Hasil Akurasi Algoritma K-Means dan K-Means + Genetic Algorithm	66

DAFTAR GAMBAR

Gambar 2. 1 Diagram Data Mining	30
Gambar 2. 2 Diagram K-Means	31
Gambar 2. 3 Diagram Algoritma Genetika	34
Gambar 3. 1 Gambar Diagram Alir Penelitian.....	39
Gambar 3. 2 Diagram Alir Modeling	41
Gambar 4. 1 Histogram Likes	49
Gambar 4. 2 Histogram Rating	49
Gambar 4. 3 Histogram Subscribers	50
Gambar 4. 4 Boxplot Likes	51
Gambar 4. 5 Boxplot Rating	51
Gambar 4. 6 Boxplot Subscribers	52
Gambar 4. 7 Centroid K-Means $K = 2$	54
Gambar 4. 8 Grafik Silhouette Score	55
Gambar 4. 9 Silhouette 2 Klaster	55
Gambar 4. 10 Centroid K-Means $K = 3$	57
Gambar 4. 11 Silhouette 3 Klaster	58
Gambar 4. 12 Titik Centroid K-Means & Genetic Algorithm $K = 2$	60
Gambar 4. 13 Titik Centroid K-Means & Genetic Algorithm $K = 3$	62

DAFTAR LAMPIRAN

Lampiran 1 Code Exploratory Data Analysis	74
Lampiran 2 Code Pre-Processing	75
Lampiran 3 Code Silhouette.....	77
Lampiran 4 Code K-Means dengan $K = 2$	79
Lampiran 5 Code K-Means dengan $K = 3$	80
Lampiran 6 Code K-Means + Genetic Algorithm $K = 2$	81
Lampiran 7 Code K-Means + Genetic Algorithm $K = 3$	84
Lampiran 8 Code Davies-Bouldin Index	87
Lampiran 9 Code Analisis Hasil Kluster	88