

TUGAS AKHIR

**OPTIMASI *HEATMAP* MENGGUNAKAN *K-PROTOTYPE*
CLUSTERING DALAM PENGELOMPOKAN DATA
KASUS PEMBUNUHAN BERANTAI**

(Studi Kasus: Pembunuhan di USA Tahun 1980-2014)



ANANDA MUHAMMAD AL-AZIZ

15102048

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

2019

TUGAS AKHIR

**OPTIMASI *HEATMAP* MENGGUNAKAN *K-PROTOTYPE*
CLUSTERING DALAM PENGELOMPOKAN DATA
KASUS PEMBUNUHAN BERANTAI**

***HEATMAP OPTIMIZATION USING K-PROTOTYPE
CLUSTERING IN THE DATA GROUPING SERIAL
HOMICIDE CASE***

(Studi Kasus: Pembunuhan di USA Tahun 1980-2014)

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



ANANDA MUHAMMAD AL-AZIZ

15102048

**PROGRAM STUDI S1 INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

2019

Lembar Pengesahan Pembimbing

**OPTIMASI *HEATMAP* MENGGUNAKAN *K-PROTOTYPE*
CLUSTERING DALAM PENGELOMPOKAN DATA
KASUS PEMBUNUHAN BERANTAI**

***HEATMAP OPTIMIZATION USING K-PROTOTYPE
CLUSTERING IN THE DATA GROUPING SERIAL
HOMICIDE CASE***

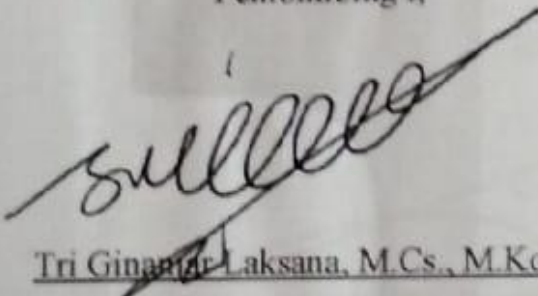
Dipersiapkan dan Disusun Oleh

**ANANDA MUHAMMAD AL-AZIZ
15102048**

Tugas Akhir Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir
Pada Hari Kamis, 15 Agustus 2019

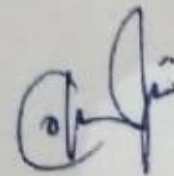
Pembimbing I,

Pembimbing II,



Tri Ginandjar Laksana, M.Cs., M.Kom.

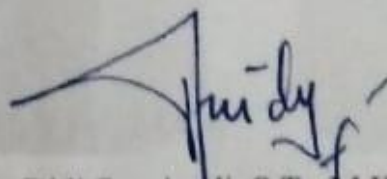
NIDN. 0407088502



Citra Wiguna, S. Kom., M. Kom.

NIDN. 0602129101

Tugas Akhir ini diterima sebagai salah satu persyaratan untuk memperoleh gelar
Sarjana Komputer
Tanggal 15 Agustus 2019
Dekan



Didi Supriyadi, S.T., M.Kom.

NIK. 13840016

Lembar Penetapan Penguji

**OPTIMASI *HEATMAP* MENGGUNAKAN *K-PROTOTYPE*
CLUSTERING DALAM PENGELOMPOKAN DATA
KASUS PEMBUNUHAN BERANTAI**

***HEATMAP OPTIMIZATION USING K-PROTOTYPE
CLUSTERING IN THE DATA GROUPING SERIAL***

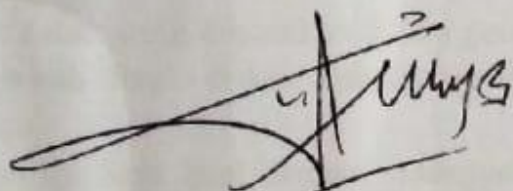
HOMICIDE CASE

Dipersiapkan dan Disusun Oleh

**ANANDA MUHAMMAD AL-AZIZ
15102048**

**Tugas Akhir Telah diuji dan Dinilai Panitia Penguji Program Studi Teknik
Informatika Fakultas Teknologi Industri dan Informatika
Institut Teknologi Telkom Purwokerto
Pada Tanggal: 15 Agustus 2019**

Ketua Penguji



(Auliya Burhanuddin, S.Si., M. Kom)

NIDN. 0630058202

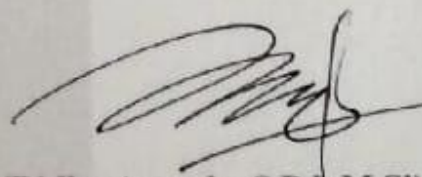
Anggota Penguji I



(Amalia Beladinna Arifa, S.Pd., M.Cs)

NIDN. 0606019201

Anggota Penguji II



(Ridho Ananda, S.Pd, M.Si)

NIDN. 0626049003

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama Mahasiswa : Ananda Muhammad Al-Aziz
NIM : 15102048
Program Studi : S1 Informatika

Menyatakan bahwa Tugas Akhir dengan judul berikut:

Optimasi *Heatmap* Menggunakan *K-Prototype Clustering* Dalam Pengelompokan Data Kasus Pembunuhan Berantai

Dosen Pembimbing Utama : Tri Ginanjar Laksana, M.Cs., M.Kom.
Dosen Pembimbing Pendamping : Citra Wiguna, S. Kom., M. Kom.

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, 15 Agustus 2019,

Yang Menyatakan.



METERAI
TEMPEL
5000
Rp5000
RUMAH KEMAHKAMAN

(Ananda Muhammad Al-Aziz)

KATA PENGANTAR

Puji syukur peneliti panjatkan kehadirat Allah S.W.T atas karunia-Nya sehingga peneliti dapat menyelesaikan penyusunan Tugas Akhir dengan lancar. Tujuan dari Tugas Akhir ini adalah memberikan kesempatan pada mahasiswa untuk melakukan studi kasus pada bidang yang mereka kuasai sehingga mahasiswa mampu menerapkan teori yang didapat selama perkuliahan sebagai penerapan ilmu pengetahuan serta pengalaman bagi mahasiswa dalam menyelesaikan suatu masalah dengan metode *computing*.

Selama penyusunan Proposal Tugas Akhir, peneliti banyak dibantu oleh beberapa pihak. Oleh karena itu peneliti ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Ali Rokhman M. Si selaku Rektor Institut Teknologi Telkom Purwokerto.
2. Bapak Didi Supriyadi S.T., M. Kom selaku Dekan Fakultas Teknologi Industri dan Informatika Institut Teknologi Telkom Purwokerto.
3. Bapak Fahrudin Mukti Wibowo S.Kom.,M.Eng., selaku Ketua Program Studi S1 Informatika Institut Teknologi Telkom Purwokerto.
4. Bapak Tri Ginanjar Laksana, M.CS, M.Kom, M. Si., selaku Pembimbing I dan Dosen di Institut Teknologi Telkom Purwokerto.
5. Ibu Citra Wiguna, S. Kom., M. Kom., selaku Pembimbing II dan Dosen di Institut Teknologi Telkom Purwokerto.
6. Dessy Karina Dewi., S.pd, Gusti Agil Saputra, dan Nova Setya Putra Utama selaku teman seperjuangan yang juga memberikan dukungan pada penulis.

Peneliti menyadari keterbatasan kemampuan dan pengetahuan yang dimiliki dalam penyusunan Proposal Tugas Akhir. Oleh karena itu, peneliti memohon maaf apabila terdapat kesalahan dalam penyusunan proposal ini. Semoga skripsi ini bermanfaat bagi peneliti dan juga teman-teman mahasiswa.

Purwokerto, 15 Agustus 2019

Hormat saya,

Ananda Muhammad Al-Aziz

DAFTAR ISI

Lembar Pengesahan Pembimbing	iii
Lembar Penetapan Penguji	iv
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN	xii
ABSTRAK	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Sebelumnya	5
2.2 Dasar Teori.....	14
2.2.1 <i>Data mining</i>	14
2.2.2 <i>Clustering</i>	15
2.2.3 <i>Silhouette Coefficient</i>	19
2.2.4 <i>Dimensionality Reduction</i>	21
2.2.5 <i>One-Hot Encoding</i>	23
2.2.6 <i>Geographic Profiling</i>	24
2.2.7 <i>Heatmap</i>	24

2.2.8 Bahasa Pemrograman Python	26
2.2.9 Bahasa Pemrograman R	26
2.2.10 Data <i>Sampling</i>	27
BAB III METODE PENELITIAN.....	30
3.1 Tahap Penelitian.....	30
3.2 Studi Literatur	31
3.3 Pengumpulan Data	31
3.4 Preprocessing Data.....	32
3.5 Seleksi Fitur	32
3.6 K-Prototype Clustering	34
3.7 Pembuatan Heatmap.....	35
3.8 Evaluasi	35
BAB IV HASIL DAN PEMBAHASAN	37
4.1 Hasil Penelitian	37
4.1.1 Pengambilan Sampel Data Acak.....	37
4.1.2 <i>Overview</i> pada Data	37
4.1.3 Preprocessing	45
4.1.4 Seleksi Fitur	48
4.1.5 Pembuatan K-Prototype Clustering	50
4.1.6 Pembuatan <i>Heatmap</i>	51
4.1.7 Evaluasi	53
4.2 Analisis Penelitian.....	54
BAB V KESIMPULAN DAN SARAN.....	56
5.1. Kesimpulan.....	56
5.2. Saran	56
DAFTAR PUSTAKA	58

DAFTAR GAMBAR

Gambar 1.1	<i>Homicide Clearance Rate</i> Keseluruhan (1965 - 2017).....	2
Gambar 2.1	<i>Circle Hypothesis: Commuter Model</i>	5
Gambar 2.2	<i>Circle Hypothesis: Marauder Model</i>	6
Gambar 2.3	Grafik Persentasi Sampel dari <i>QRange</i>	7
Gambar 2.4	Jarak Rata-Rata Galat dan Standar Deviasi dari Dua Grup dalam Dua Tahap dan Galat dari Dragnet	9
Gambar 2.5	<i>Silhouette Coefficient</i> $a(i)$ dan (b) dengan $a(i)$ sebagai Rata-Rata <i>Similarity</i> i ke Semua Objek a	20
Gambar 2.6	Konversi Data Mentah ke Bentuk Matriks Indikator (<i>Dummy Variable</i>)	22
Gambar 2.7	Sampel Data yang Digunakan pada Sumber	25
Gambar 2.8	Matriks A sebagai Data, Matriks A^T sebagai Data Tranpose, dan B Hasil <i>Dot Product</i>	25
Gambar 2.9	Ukuran Sampel Berbanding Total Populasi pada Nilai <i>Confidence</i> = 95%.....	29
Gambar 3.1	Diagram Alir Penelitian	30
Gambar 3.2	Diagram Alir <i>K-Prototype Clustering</i>	34
Gambar 3.3	Diagram Alir Pembuatan <i>Heatmap</i>	35
Gambar 4.1	Memisahkan Jumlah Kasus ke Setiap Negara Bagian	38
Gambar 4.2	<i>Data Frame</i> Longitude dan Latitude Negara Bagian.....	38
Gambar 4.3	<i>Data Frame</i> Hasil Penggabungan Negara Bagian, Jumlah Pembunuhan, serta <i>Centroid Longitude</i> dan <i>Centroid Latitude</i>	39
Gambar 4.4	Los Angeles Menduduki Jumlah Kasus Pembunuhan Tertinggi di California	40
Gambar 4.5	Gambar Peta Setelah Ditambahkan Label dan Penyesuaian Style (Sampel).....	41
Gambar 4.6	Gambar Peta Setelah Ditambahkan Label dan Penyesuaian Style (Data Asli).....	(42)

Gambar 4.7 Korelasi Antar Atribut	43
Gambar 4.8 Korelasi Antar Atribut (Setelah <i>Crime Type</i> Dihapus).....	44
Gambar 4.9 Hasil Cetak Fungsi “np.unique” pada Data Numerik	46
Gambar 4.10 Hasil Cetak “np.unique” pada Data <i>Categorical Month, CrimeSolved,</i> dan <i>Relationship</i>	47
Gambar 4.11 Hasil “np.unique” pada Data <i>Categorical Weapon</i> , dan <i>City</i>	47
Gambar 4.12 Hasil “np.unique” pada Data <i>Categorical</i> Karakteristik Korban....	47
Gambar 4.13 Hasil “np.unique” pada Data <i>Categorical</i> Karakteristik Pelaku.....	48
Gambar 4.14 <i>Spike</i> pada Korban Pembunuhan Diusia 99 Tahun	48
Gambar 4.15 Hasil Interpretasi <i>PCA</i> pada Data Pembunuhan.....	49
Gambar 4.16 Hasil Cetak “np.unique” pada Atribut Insiden.....	50
Gambar 4.17 Hasil Fungsi <i>dimdesc</i> dari <i>MCA</i>	50
Gambar 4.18 Bentuk Objek <i>K-Prototype Clustering</i>	51
Gambar 4.19 <i>Dataframe</i> yang Sudah Diberi Label dengan <i>Cluster</i>	51
Gambar 4.20 <i>Heatmap 2D</i> pada Seleksi Fitur Heuristik (tanpa <i>cluster</i>).....	52
Gambar 4.21 <i>Heatmap 2D</i> pada Seleksi Fitur <i>PCA</i> dan <i>MCA</i> (dengan <i>cluster</i>)...52	
Gambar 4.22 <i>Heatmap 2D</i> pada Seleksi Fitur Heuristik (dengan <i>cluster</i>).....52	

DAFTAR TABEL

Tabel 2.1	Kriteria Lokasi Kejahatan Berdasarkan <i>Bottom-up</i>	12
Tabel 2.2	Tabel Penelitian Sebelumnya (<i>State of The Art</i>).....	12
Tabel 3.1	Atribut pada Dataset Kasus Pembunuhan di Amerika Serikat Tahun 1980 - 2014.....	32
Tabel 4.1	Hasil Perhitungan <i>Cochran Formula</i>	37
Tabel 4.2	Informasi Korelasi Antar Atribut (Di atas 0,4 dan Dibawah -0,4)	44
Tabel 4.3	Variabel yang Diambil Berdasarkan Pendekatan Heuristik.....	49
Tabel 4.4	Skor <i>Silhouette Coefficient</i> dari Seleksi Data Secara Heuristik dari <i>Cluster K = 2</i> hingga <i>K = 10</i>	53
Tabel 4.5	Skor <i>Silhouette Coefficient</i> dari Seleksi Data <i>PCA</i> dan <i>MCA</i> dari <i>Cluster K = 2</i> hingga <i>K = 10</i>	53

DAFTAR LAMPIRAN

Lampiran 1. Sampel Dataset Pembunuhan di Amerika Serikat Tahun (1965-2014)

Lampiran 2. *Source Code K-Prototype* dan *Silhouette Index*

Lampiran 3. Source Code Pembuatan Heatmap