



**IMPLEMENTASI DATA MINING UNTUK *CLUSTERING* PENDUDUK
DI KECAMATAN JIRAK JAYA MENGGUNAKAN ALGORITMA
*RANDOM FOREST***

KARYA AKHIR

DINA PRATAMI PUTRI

211420056

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA PALEMBANG
TAHUN 2025**

HALAMAN PENGESAHAN

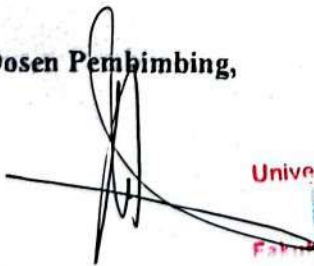
IMPLEMENTASI DATA MINING UNTUK *CLUSTERING* PENDUDUK
DI KECAMATAN JIRAK JAYA MENGGUNAKAN ALGORITMA
RANDOM FOREST

DINA PRATAMI PUTRI

211420056

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer (S.Kom) pada Program Studi Teknik Informatika

Dosen Pembimbing,



Alek Wijaya, S.Kom. MIT.

Palembang, 26 - Agustus - 2025
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,






Dr. Tata Sutabri, S.Kom., MMSI., MKM

HALAMAN PERSETUJUAN

Skripsi berjudul "Implementasi Data Mining Untuk *Clustering* penduduk Di Kecamatan Jirak Jaya Menggunakan Algoritma *Random Forest*" oleh "Dina Pratami Putri", Telah dipertahankan di depan komisi penguji pada hari Jum'at tanggal 15 Agustus 2025.

Komisi penguji

1. Ketua : Alek Wijaya, S.Kom. MIT.  (.....)
2. Anggota : Nurul Huda, M.Kom  (.....)
3. Anggota : Nurul Adha Oktarini Sapitri, M.Kom  (.....)

Mengetahui,
Program Studi Teknik Informatika
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua Program Studi,

Universitas Bina Darma
Fakultas Sains Teknologi

Alek Wijaya, S.Kom. MIT

HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Dina Pratami Putri

Nim : 211420056

Dengan ini menyatakan bahwa:

1. Karya tulis saya adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan;
4. Saya bersedia karya tulis ini di cek keasliannya menggunakan plagiarism checker serta di unggah di internet, sehingga dapat diakses secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidak benaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 13-8-2025
ng membuat surat pernyataan



Dina Pratami Putri

NIM: 211420056

MOTTO DAN PERSEMBAHAN

MOTTO

"Data adalah harta karun masa depan dengan kecerdasan buatan, kita membuka pintu kesejahteraan bagi masyarakat."

PERSEMBAHAN

Tugas Akhir Skripsi ini saya persembahkan untuk :

1. Kedua orang tua tercinta, yang selalu menjadi pilar kekuatan dan doa utama dalam setiap langkah perjuangan saya. Terima kasih atas kasih sayang, pengorbanan, dan dukungan tak terbatas yang telah diberikan.
2. Dosen pembimbing dan Dosen penguji yang telah membimbing, mengarahkan, dan memberikan ilmu pengetahuan berharga dalam bidang data mining dan algoritma machine learning.
3. Sahabat dan keluarga, yang senantiasa menjadi motivasi dan penyemangat di tengah tantangan penelitian ini.
4. Semoga karya ini bermanfaat bagi kemajuan ilmu pengetahuan, khususnya dalam implementasi data mining untuk clustering penduduk di Kecamatan Jirak Jaya, serta dapat memberikan kontribusi positif bagi masyarakat dan pembangunan daerah.

ABSTRAK

Struktur sosial masyarakat, terutama status sosial, memainkan peran penting dalam akses terhadap sumber daya dan kualitas hidup penduduk. Di Indonesia, data Badan Pusat Statistik (BPS) tahun 2020 menunjukkan bahwa 10,19% penduduk masih hidup di bawah garis kemiskinan, dengan kesenjangan signifikan antara masyarakat perkotaan dan pedesaan serta latar belakang pendidikan dan pekerjaan. Stratifikasi sosial, yang telah ada sejak zaman kuno, membagi masyarakat ke dalam kelas kaya, menengah, dan miskin berdasarkan status sosial yang dapat bersifat ascribed, achieved, atau assigned. Pada umumnya didalam setiap negara selau terdapat tiga unsur, yakni orang-orang kaya, orang-orang miskin, dan orang-orang menengah. Pelapisan ekonomi dapat dilihat dari segi pendapatan, kekayaan, dan pekerjaan. Dalam konteks ekonomi, stratifikasi terlihat dari pendapatan, kekayaan, dan pekerjaan, di mana orang-orang dengan kapasitas ekonomi tinggi menempati lapisan atas. Tujuan dari penelitian ini untuk *clustering* penduduk di Kecamatan Jirak Jaya. Dalam penelitian ini menggunakan metode *Random Forest* variabel yang digunakan dalam penerapan metode *Random Forest* ada dua yaitu variabel independen dan variabel dependen. Metode *Random Forest* dipilih sebagai teknik klasifikasi yang efektif untuk menangani data besar dan tidak lengkap. Fokus penelitian ini adalah pada Kecamatan Jirak Jaya, Kabupaten Musi Banyuasin, yang mayoritas penduduknya berprofesi sebagai petani sawit dan karet. Penelitian ini bertujuan untuk melakukan *clustering* penduduk berdasarkan status sosial dan penghasilan, sehingga dapat menjadi acuan dalam penyaluran bantuan pemerintah. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan yang berguna dalam memahami stratifikasi sosial dan ekonomi di masyarakat. Hasil penelitian ini cluster rendah menunjukkan nilai test 56.67%, cluster sedang 40% dan cluster tinggi 3,33%. Dimana angka ini menunjukkan bahwa lebih dari separuh Masyarakat Kec. Jirak Jaya berada di cluster rendah dan hanya beberapa persen saja di cluster tinggi.

Kata Kunci : *Status sosial, Kesenjangan ekonomi, Penghasilan, Pengeluaran dan Jenis Pekerjaan.*

KATA PENGANTAR

Puji Syukur penulis ucapkan kehadiran Allah SWT karena berkat rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Implementasi Data Mining Untuk *Clustering* penduduk Di Kecamatan Jirak Jaya Menggunakan Algoritma *Random Forest*” ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana komputer pada Fakultas Sains Teknologi Program Studi Teknik Informatika Universitas Bina Darma Palembang. Penulis menyadari bahwa dalam penyusunan skripsi ini terwujud berkat bantuan, arahan, bimbingan dan doa dari berbagai pihak. Oleh karena itu penulis menyampaikan terimakasih kepada:

1. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma.
2. Dr. Tata Sutabri, S.Kom, MMSI, MKM selaku Dekan Fakultas Sains Teknologi Universitas Bina Darma.
3. Alek Wijaya, S.Kom., M.I.T. selaku Ketua Program studi Teknik Informatika Universitas Bina Darma, sekaligus Dosen Pembimbing karya akhir yang telah memberikan bimbingan dan arahan dalam menyelesaikan skripsi ini.
4. Dosen fakultas sains teknologi Universitas Bina Darma yang tidak dapat sebut satu persatu karena telah banyak memberikan banyak ilmu dan masukan kepada penulis.
5. Seluruh Staff Kecamatan Jirak yang telah memberikan bimbingan serta arahan selama penelitian dan masyarakat Jirak Jaya yang telah terlibat dalam penelitian ini.
6. Kedua orang tua, Ayah Amin Budiono dan ibunda Sulastri. Dan kedua Adik penulis, Novi Zahra Putri dan Fadlan Rafa Saputra atas segala pengorbanan, doa, semangat dan kasih sayang, serta perhatian dan dukungan dalam penulisan skripsi ini sehingga penulis mampu menyelesaikan karya akhir ini dengan baik.
7. Kepada sahabat-sahabatku yang takalah penting kehadirannya Novitria Indriany, A.Md.Keb, Ratmi Sartika, S.Pd, Syazwana xena indah rahmadhani safitri, Nopi Pitriana, Yosandra Saputri dan teman-teman Angkatan 2021 Teknik Informatika Universitas Bina Darma.

8. Kepada seseorang yang tak kalah penting kehadirannya Aldi Ansyah, S.E. telah banyak menjadi bagian dari perjalanan hidup penulis. Berkontribusi banyak dalam penulisan karya akhir ini, baik tenaga, materi, maupun waktu kepada penulis. Sudah selalu mendukung, mendoakan, menghibur, mendengar keluh kesah, sehingga penulis dapat menyelesaikan karya akhir ini.

Penulis menyadari bahwa dalam penyusunan karya akhir ini masih terdapat banyak kekurangan, oleh karena itu penulis mengharapkan kritik dan saran yang membangun untuk perbaikan dalam penulisan karya akhir selanjutnya.

Palembang, 15 Agustus 2025



Dina Pratami Putri

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERSETUJUAN	ii
HALAMAN PERNYATAAN	iii
MOTO DAN PERSEMBAHAN	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	
2.1 Data Mining	7
2.2 <i>Random Forest</i>	11
2.2.1 Karakteristik <i>Random Forest</i>	12
2.2.2 Proses Kerja <i>Random Forest</i>	14
2.3 <i>Clustering</i> Penduduk.....	15
2.4 Kecamatan Jirak Jaya.....	16
2.5 Penelitian terdahulu yang Relevan.....	18
2.6 Kerangka Pemikiran.....	19
BAB III METODE PENELITIAN	
3.1 Pengumpulan Data	21
3.1.1 Data Primer	21
3.1.2 Data Sekunder	22
3.1.3 Variabel Penelitian	22
3.2 Pra-Pemrosesan Data.....	24
3.2.1 Normalisasi kolom dan Penyesuaian Nama.....	24
3.2.2 Penghapusan Data Duplikat	25
3.2.3 Penanganan <i>Messing Velues</i>	25
3.2.4 <i>Encoding</i> Variabel Kategorial	26
3.2.5 Transformasi Data Numerik dan Rentang.....	27
3.2.6 Ekstrasi Nilai Jumlah Tanggungan.....	27
3.2.7 Pembersihan Kolom NIK.....	28
3.2.8 Penyusunan Ulang Kolom dan Penyimpanan <i>Output</i>	28

3.3 Pemilihan Fitur.....	29
3.3.1 Fitur-fitur Terpilih	30
3.3.2 Alasan Pemelihan Fitur	31
3.4 Penerapan Algoritma <i>Random Forest</i>	33
3.4.1 Analisis Deskriptif.....	34
3.4.2 Permodelan <i>Random Forest</i>	40
3.4.3 <i>Tuning</i> Parameter	44

BAB IV HASIL DAN PEMBAHASAN

4. 1 Pengujian dan Hasil	46
4.1.1 Pembagian Data <i>Training</i> dan Data <i>Testing</i>	46
4.1.2 Analisis Klasifikasi <i>Random Forest</i>	47
4. 2 Analisis Hasil <i>Cluster</i>	51
4.2.1 Visualisasi dan Interpretasi.....	52
4. 3 Evaluasi Model	60
4.3.1 Akurasi Model	61
4.3.2 <i>Confusion Matrix</i>	62
4.3.3 <i>Classification Report</i>	63
4.3.4 Kesimpulan evaluasi Model	65

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan	67
5.2 Saran.....	68

DAFTAR PUSTAKA

LAMPIRAN-LAMPIRAN

DAFTAR GAMBAR

Gambar 2.1	Visualisasi cara kerja <i>Random Forest</i>	12
Gambar 3.1	Laporan Hasil Pre-Processing Data	29
Gambar 4.1	Contoh salah satu pohon Keputusan model <i>Clustering</i>	48
Gambar 4.2	Barchart Distribusi Jumlah Penduduk per <i>Cluster</i>	52
Gambar 4.3	Doughnut Charth persentase Distribusi <i>Cluster</i>	53
Gambar 4.4	<i>Features Imfortance</i> metode <i>Random Forest</i>	55
Gambar 4.5	Tabel Data <i>Cluster</i> penduduk	57
Gambar 4.6	Tabel Data <i>Cluster</i> Penduduk Hasil Filter Rendah.....	59
Gambar 4.7	Tabel Data <i>Cluster</i> Penduduk Hasil Filter Sedang	60
Gambar 4.8	Tabel Data <i>Cluster</i> Penduduk Hasil Filter Tinggi.....	60

DAFTAR TABEL

Tabel 2.1	Proses Data Mining <i>Clustering</i> Penduduk.....	9
Tabel 2.2	Jumlah Penduduk Kecamatan Jirak Jaya 2020-2024.....	18
Tabel 3.1	Variabel Penelitian yang digunakan.....	23
Tabel 3.2	Perubahan kolom setelah dinormalisasi.....	24
Tabel 3.3	Penanganan Duplikat Data.....	25
Tabel 3.4	Penanganan <i>Missing Values</i>	25
Tabel 3.5	<i>Encoding</i> kolom Pendidikan	26
Tabel 3.6	<i>Encoding</i> kolom Pekerjaan	26
Tabel 3.7	<i>Encoding</i> kolom Usia.....	27
Tabel 3.8	Transformasi Data <i>Numerik</i> dan Rentang.....	27
Tabel 3.9	Ekstrasi Nilai Jumlah Tanggungan	28
Tabel 3.10	Pembersihan kolom NIK	28
Tabel 3.11	Fitur Data Terpilih.....	30
Tabel 3.12	Konversi Nilai Nominal.....	31
Tabel 3.13	Nilai Rasio Pengeluaran-Pendapatan.....	32
Tabel 3.14	Rasio Tabungan Per tanggungan.....	32
Tabel 3.15	<i>Encoding</i> Fitur Usia-Pendidikan.....	32
Tabel 3.16	Distribusi Fitur Pendidikan	34
Tabel 3.17	Distribusi Fitur Pekerjaan	35
Tabel 3.18	Distribusi Fitur jumlah tanggungan	35
Tabel 3.19	Distribusi Fitur jumlah pengeluaran bulanan.....	36
Tabel 3.20	Distribusi Fitur Tabungan bulanan.....	36
Tabel 3.21	Distribusi Fitur usia kelompok	37
Tabel 3.22	Distribusi Fitur Rasio Pengeluaran-Pendapatan	37
Tabel 3.23	Distribusi Fitur Tabungan per tanggungan	38
Tabel 3.24	Distribusi Fitur total Tanggungan	39
Tabel 3.25	Distribusi Fitur total pendapatan.....	39
Tabel 3.26	Indikator Nilai penentu <i>Cluster</i> per fitur	40
Tabel 3.27	Data Latih	41

Tabel 3.28	Hasil <i>Tuning</i> Parameter Metode <i>Random Forest</i>	45
Tabel 4.1	Proporsi Data <i>Training</i> dan Data <i>Testing</i>	47
Tabel 4.2	Tabel Distribusi <i>Cluster</i>	55
Tabel 4.3	Nilai <i>Feature Importance</i> metode <i>Random Forest</i>	56
Tabel 4.4	Contoh Perbandingan Data <i>Cluster</i> Asumsi dan Prediksi Model.	58
Tabel 4.5	Hasil Akurasi <i>Random Forest</i>	58
Tabel 4.6	Confusion Matrix dari Data Testing <i>Random Forest</i>	59
Tabel 4.7	<i>Classification Report</i> Model <i>Random Forest</i>	62

