

**PEMETAAN PREVALENSI DAN ANALISIS FAKTOR PENYAKIT
TUBERCULOSIS DI INDONESIA MENGGUNAKAN K-MEANS
CLUSTERING DAN SUPPORT VECTOR REGRESSOR**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh

Fadhila Rahma Dwicahyani

217006516053



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2024**

**PEMETAAN PREVALENSI DAN ANALISIS FAKTOR PENYAKIT
TUBERCULOSIS DI INDONESIA MENGGUNAKAN K-MEANS
CLUSTERING DAN SUPPORT VECTOR REGRESSOR**

SKRIPSI SARJANA SISTEM INFORMASI

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Fadhila Rahma Dwicahyani

217006516053



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2024**

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

Pemetaan Prevalensi Dan Analisis Faktor Penyakit Tuberculosis Di Indonesia Menggunakan K-Means Clustering Dan Support Vector Regressor

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 26 Februari 2025



Pembimbing 1

Rima Tamara Aldisa, S.Kom., M.Kom

NIDN.0318019401

Ketua Program Studi



Dr. Andrianingsih, S.Kom, MMSI

NIDN.0303097902

LEMBAR PERSETUJUAN JUDUL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama	: Fadhila Rahma Dwicahyani
NPM	: 217006516053
Fakultas/Akademi	: Fakultas Teknologi Komunikasi dan Informatika
Program Studi	: Sistem Informasi
Tanggal Sidang	: 26 Maret 2025

JUDUL DALAM BAHASA INDONESIA :

Pemetaan Prevalensi Dan Analisis Faktor Penyakit Tuberculosis Di Indonesia Menggunakan K-Means Clustering Dan Support Vector Regressor

JUDUL DALAM BAHASA INGGRIS :

Prevalence Mapping and Factor Analysis of Tuberculosis Disease in Indonesia Using K-Means Clustering and Support Vector Regressor

TANDA TANGAN DAN TANGGAL		
Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 4 Maret 2025	TGL : 4 Maret 2025	TGL : 4 Maret 2025
 Rima	 Dr. Rada Nugroho, M.M.G.	 fadhila Rahma Dwicahyani

HALAMAN PERNYATAAN ORISINALITAS

HALAMAN PERNYATAAN ORISINALITAS

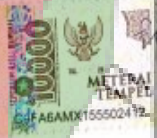
Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

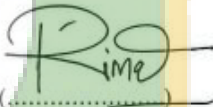
Nama : Fadhila Rahma Dwicahyani

NIM : 217006516053

Tanda Tangan : 

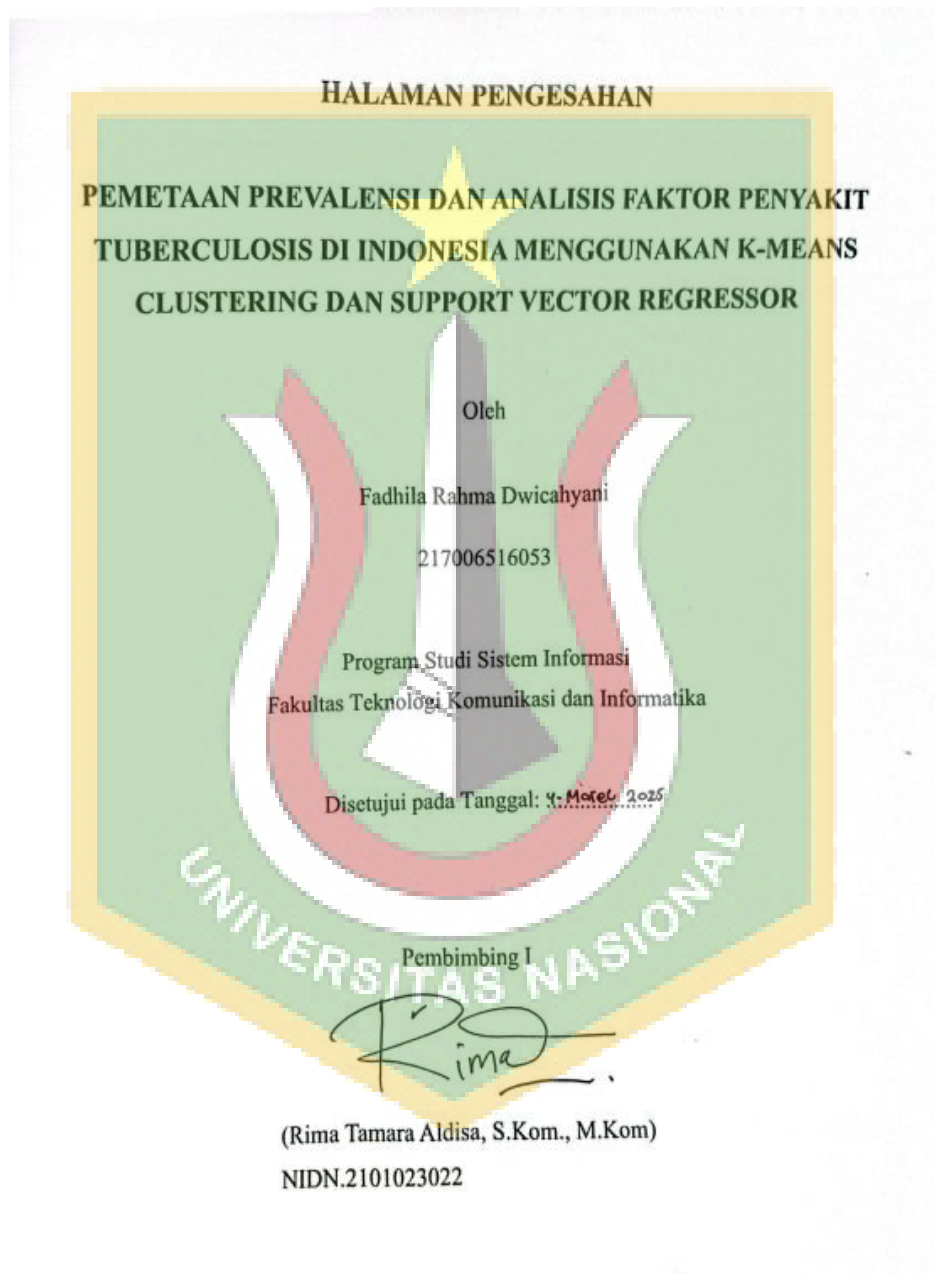
Tanggal :  4 Maret 2025

Mengetahui

Pembimbing I : Rima Tamara Aldisa, S.Kom., M.Kom. ()

UNIVERSITAS NASIONAL

HALAMAN PENGESAHAN



HALAMAN PENGESAHAN



KATA PENGANTAR

Puji syukur penulis sampaikan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, yang telah memungkinkan penulis menyelesaikan skripsi berjudul "Pemetaan Prevalensi Dan Analisis Faktor Penyakit Tuberculosis Di Indonesia Menggunakan K-Means Clustering dan Support Vector Regressor" sebagai persyaratan kelulusan Program Studi Sarjana Sistem Informasi di Fakultas Teknologi Komunikasi dan Informatika. Penyelesaian penelitian dan penyusunan skripsi ini tidak lepas dari dukungan berbagai pihak, oleh karena itu, penulis ingin menyampaikan penghargaan kepada dosen pembimbing Tugas Akhir atas dedikasi, bimbingan, motivasi, dan pengertian yang diberikan selama proses penelitian dan penulisan skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Ibu dari penulis, kakak dari penulis, dan tidak lupa untuk alm. ayah dari penulis, yang selalu memberikan bantuan dan dukungan yang telah diberikan selama berjalannya proses penelitian ini hingga penyelesaian dari studi ini.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom. selaku Dekan Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional yang telah memberikan izin bagi penulis untuk menulis skripsi ini.
3. Ibu Dr. Andrianingsih, S.Kom., MMSI. selaku kepala program studi sistem informasi.
4. Ibu Rima Tamara Aldisa, S.Kom., M.Kom. selaku dosen pembimbing penulis yang telah memberikan ilmu yang bermanfaat, serta memberikan arahan juga bimbingan selama proses penelitian sampai akhir penulisan skripsi.
5. Seluruh dosen dan seluruh staff di Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional yang telah memberikan ilmu bermanfaat serta memberikan bantuan selama berjalan nya proses studi ini.

6. Sahabat penulis; Farda Nur Maulida, Sambitha Ayeshanalla Asmoro, dan Vilena Aliffa yang selalu mendukung, serta memberikan bantuan, doa juga motivasi kepada penulis.
7. Grup Dunia Per-sad an (Abija Bohdan Prasetio, Achmad Maulana Alawi, Arga Mahesa Putra, Aqsal Rizky Rafsanjani Camelia Nur Hidayah, Rizky Wahyu Sugandhi, Ronald Lakhomizaro Gulo, Tegar Suryatama, Sekar Keisya Meidy Santoso) selaku orang terdekat selama masa perkuliahan dan penelitian memberikan banyak bantuan. Dan grup Penghuni Sanhok (Ilyasa Aufa, Nayna Sonya P.) selaku orang terdekat di luar masa studi ini yang bersedia untuk bertukar pikiran, dan juga memberikan motivasi selama proses penelitian ini berjalan.
8. Serta seluruh pihak yang telah membantu, namun tidak bisa disebutkan satu per satu.

Penulis mengakhiri kata pengantar ini dengan harapan agar Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan keberkahan. Penulis juga terbuka untuk menerima kritik dan saran yang membangun, sembari berharap skripsi ini dapat memberikan kontribusi positif di bidang Teknologi Informatika.

ABSTRAK

Kuman *Mycobacterium Tuberculosis* menjangkiti bagian paru-paru, menyebabkan penyakit menular yang disebut TBC. Demam, kurus, dan batuk adalah gejala TBC yang mengerikan yang membunuh orang dewasa dan anak-anak. Penyakit TBC paru secara umum disebabkan oleh beberapa faktor, seperti lingkungan, cuaca, tempat tinggal, dan keturunan. Saat ini, tuberkulosis (TBC) masih menjadi masalah bagi kesehatan masyarakat Indonesia. Sebuah laporan dari Organisasi Kesehatan Dunia (WHO) menyatakan bahwa Indonesia berada di antara negara dengan tingkat infeksi TBC tertinggi di dunia. Hasil Algoritma *K-Means Clustering* menunjukkan bahwa TBC lebih umum di daerah dengan kepadatan penduduk tinggi dan akses layanan kesehatan terbatas, dengan Cluster 3 memiliki proporsi kasus terbesar (38,91%). Tren tahunan yang berbeda menunjukkan bahwa ada perlunya pemantauan lebih lanjut. evaluasi model *Support Vector Regressor* (SVR) menunjukkan hasil akurasi prediksi masih rendah yaitu dengan nilai MAE 68,41, RMSE 76,37, dan R^2 -0,03, menunjukkan bahwa model mengalami underfitting.

Kata Kunci: Tuberculosis, *K-Means Clustering*, *Support Vector Regressor*, Pengelompokan, Regresi, Prediksi, Visualisasi.

ABSTRACT

The Mycobacterium Tuberculosis germ infects the lungs, causing an infectious disease called tuberculosis. Fever, emaciation and coughing are terrible symptoms of TB that kill adults and children. Pulmonary tuberculosis disease is generally caused by several factors, such as environment, weather, residence, and heredity. Today, tuberculosis (TB) is still a problem for Indonesian public health. A report from the World Health Organization (WHO) states that Indonesia is among the countries with the highest TB infection rates in the world. The results of the K-Means Clustering Algorithm show that TB is more common in areas with high population density and limited access to health services, with Cluster 3 having the largest proportion of cases (38.91%). The evaluation of the Support Vector Regressor (SVR) model shows that the prediction accuracy is still low with MAE 68.41, RMSE 76.37, and R^2 -0.03, indicating that the model is underfitting.

Keywords: K-Means Clustering, Pengelompokkan, Prediksi, Regresi, Support Vector Regressor, Tuberkulosis, Visualisasi.

DAFTAR ISI

KATA PENGANTAR.....	ii
ABSTRAK	v
ABSTRACT	iv
DAFTAR ISI	v
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Kontribusi	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Penelitian Terdahulu.....	8
2.2 Landasan Teori	17
2.2.1 Tuberculosis.....	17
2.2.2 Machine Learning.....	18
2.2.3 Python	18
2.2.4 Power BI	18
2.2.5 Google Colab	19
2.2.6 Streamlit	19

2.2.7 K-Means Clustering.....	19
2.2.8 Support Vector Regressor	20
BAB III METODE PENELITIAN	27
3.1 Waktu Penelitian	27
3.2 Fokus Penelitian	28
3.4 Tahapan Penelitian	29
BAB IV HASIL DAN DISKUSI.....	34
4.1 Pengambilan Data	34
4.2 Preprocessing Data	35
4.2.1 Data Cleaning.....	35
4.2.2 Transformasi Data	38
4.3.1 K-Means Clustering.....	40
4.3.2 Support Vector Regressor	50
4.4. Perhitungan Evaluasi Model.....	57
4.4.1 Mean Absolute Error (MAE).....	58
4.4.2 Mean Squared Error (MSE)	58
4.4.3 Root Mean Squared Error (RMSE).....	59
4.4.4 R-Squared (R^2)	60
4.5 Hasil Support Vector Regressor	61
4.6 Visualisasi Hasil	62
4.6.1 Visualisasi Clustering	63
4.6.2 Visualisasi Akumulasi TBC Tahun	64
4.6.3 Visualisasi Distribusi Prevalensi TBC Berdasarkan Klasterisasi	65

4.6.4 Visualisasi Populasi.....	66
4.6.5 Visualisasi Distribusi Akses Layanan Kesehatan.....	67
4.7. Hasil Streamlit	68
4.7.1 Home	68
4.7.2 Modeling	69
4.7.3 Visualisasi	70
BAB V KESIMPULAN DAN SARAN	71
5.1 Kesimpulan	71
5.2 Saran.....	72
DAFTAR PUSTAKA.....	73
LAMPIRAN.....	74



DAFTAR TABEL

Tabel 2.1 Penelitan Terdahulu	8
Tabel 3. 1 Waktu Penelitian.....	27
Tabel 4. 1 sample data	51
Tabel 4. 2 Data Setelah Dikonversi.....	51
Tabel 4. 3 Sample Dataset.....	57



DAFTAR GAMBAR

Gambar 3. 1 Flowchart Tahapan Penelitian	29
Gambar 3. 2 Flowchart Perancangan Sistem	32
Gambar 4. 1 Dataset Tuberculosis	34
Gambar 4. 2 Mengecek Duplikat Data.....	36
Gambar 4. 3 Mengecek Tipe Data.....	36
Gambar 4. 4 Mengecek Jumlah Missing Value.....	37
Gambar 4. 5 Label Encoding	38
Gambar 4. 6 Variabel K-Means Clustering.....	43
Gambar 4. 7 Elbow Method.....	44
Gambar 4. 8 Silhouette Score.....	45
Gambar 4. 9 Visualisasi Clustering Prevalensi TBC.....	46
Gambar 4. 10 Visualisasi Clustering Populasi	47
Gambar 4. 11 Total Cluster	48
Gambar 4. 12 Daerah Dengan Prevalensi Tertinggi.....	49
Gambar 4. 13 Nilai Prevalensi Tertinggi.....	49
Gambar 4. 14 Evaluasi Model SVR 80:20.....	56
Gambar 4. 15 Evaluasi Model SVR 70:30.....	56
Gambar 4. 16 Permutation Feature Importance.....	61
Gambar 4. 17 Dashboard kasus Tuberculosis	62
Gambar 4. 18 Visualisasi Clustering.....	63
Gambar 4. 19 Populasi Kepadatan Penduduk	64
Gambar 4. 20 distribusi prevalensi.....	65
Gambar 4. 21 Visualisasi Populasi.....	66
Gambar 4. 22 Distribusi Akses Layanan Kesehatan	67
Gambar 4. 23 Halaman Home.....	68
Gambar 4. 24 Halaman Modeling.....	69

Gambar 4. 25 Halaman Visualisasi 70

