

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Semua orang, termasuk pemerintah di tingkat provinsi, bertanggung jawab atas kesehatan. Pemerintah provinsi bertanggung jawab untuk mengawasi kualitas layanan kesehatan di tingkat provinsi dan memastikan bahwa layanan tersebut memenuhi standar dan pedoman nasional. Salah satu tugas Dinas Kesehatan adalah memantau kasus penyakit di wilayahnya. Menurut ketentuan Kementerian Kesehatan RI, ada sekitar tiga belas jenis penyakit yang harus dipantau dan dilaporkan secara teratur.

Penyakit menular yang disebut *Tuberkulosis* (TBC) disebabkan oleh kuman *Mycobacterium tuberculosis* yang menjangkit bagian organ paru-paru. Berbagai gejala TBC termasuk demam, kurus, dan batuk. Penyakit yang mengerikan yang membunuh orang dewasa dan anak-anak. Beberapa faktor, seperti lingkungan, cuaca, tempat tinggal, dan keturunan, secara umum bertanggung jawab atas penyakit TBC paru (Toresa,2020).

Hingga saat ini, *Tuberkulosis* (TBC) adalah salah satu penyakit menular yang masih menjadi masalah bagi kesehatan masyarakat Indonesia. Laporan yang dikeluarkan oleh Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa Indonesia termasuk dalam kelompok negara dengan tingkat infeksi TBC tertinggi di seluruh dunia. Penyakit ini tidak hanya mempengaruhi kesehatan seseorang, tetapi juga menyebabkan masalah ekonomi bagi mereka, baik secara individu maupun untuk sistem kesehatan secara keseluruhan. Di Indonesia, prevalensi TBC masih tinggi. Ini disebabkan oleh sejumlah faktor, termasuk keadaan sosial

ekonomi, akses ke layanan kesehatan, dan tingkat kepatuhan terhadap pengobatan, meskipun upaya pengendalian dan pencegahan telah dilakukan.

Dalam konteks ini, analisis data akan sangat penting untuk memahami pola prevalensi TBC di Indonesia dan faktor-faktor yang memengaruhinya. Metode berbasis data membantu menemukan daerah dengan prevalensi TBC yang tinggi. Mereka juga memberikan wawasan untuk mendukung pembuatan kebijakan dan penerapan strategi pengendalian pada penyakit *Tuberculosis* (TBC) yang lebih baik. Untuk menganalisis faktor-faktor yang berkontribusi terhadap penyebaran *Tuberculosis* dan mengelompokkan wilayah berdasarkan karakteristik tertentu, algoritma *Support Vector Regressor* (SVR) dan *K-Means Clustering* dapat digunakan. Proses ini, yang memanfaatkan pendekatan *machine learning* untuk mengolah data yang tersedia secara lebih terstruktur dan akurat, diharapkan dapat memberikan kontribusi dalam pengendalian tuberkulosis di Indonesia.

*Machine Learning* adalah Salah satu bagian dari *artificial intelligence* yang berfokus pada pemahaman pola untuk mencari solusi terbaik, memungkinkan komputer untuk secara otomatis menemukan solusi terbaik dari data (Siswoyo,2020). Dalam hal pengendalian *Tuberculosis* (TBC), machine learning dapat digunakan untuk menganalisis prevalensi *Tuberculosis* (TBC) di berbagai wilayah di Indonesia, mengelompokkan data berdasarkan karakteristik wilayah dan faktor-faktor pendukungnya, serta mengidentifikasi determinan yang memengaruhi tingkat keberhasilan penanganan *Tuberculosis* (TBC). *K-Means Clustering* dan *Support Vector Regressor* (SVR) akan digunakan untuk membantu menganalisis data secara menyeluruh, menemukan pola distribusi penyakit, dan memberikan wawasan yang mendalam untuk mendukung pengambilan keputusan strategis dalam upaya pengendalian dan penanggulangan *Tuberculosis* (TBC).

Salah satu algoritma untuk melakukan pengelompokkan data menjadi banyak kelompok dengan banyak *cluster* adalah *K-Means Clustering*. Algoritma *K-Means*

*Clustering* adalah salah satu dari banyak teknik data mining yang tersedia. Ini efektif dalam membuat cluster dari tumpukan data atau mengelompokkan data ke dalam beberapa cluster sehingga data yang berbeda menjadi bagian dari cluster yang berbeda (Minarni et al., 2021).

*Support Vector Regressor* (SVR) adalah kemajuan dari *Support Vector Machine* (SVM) dalam regresi. SVM adalah metode klasifikasi yang melihat daerah-daerah yang terpisah dari kategori yang ditemukan dalam pengamatan. Tujuan SVR adalah untuk menemukan fungsi  $f(x)$  sebagai *hyperplane*, atau garis pembagi, yang merupakan fungsi regresi yang cocok untuk semua data input dengan kesalahan yang paling rendah (Firman Akbar, Dkk, 2022).

Penelitian ini akan menggunakan kumpulan data yang diambil dari berbagai sumber terkait prevalensi dan faktor-faktor yang memengaruhi penyakit *Tuberculosis* (TBC) di Indonesia. Algoritma K-Means Clustering akan digunakan untuk mengidentifikasi wilayah dengan tingkat *Tuberculosis* (TBC) tertinggi, sementara algoritma *Support Vector Regressor* akan diterapkan untuk menganalisis faktor-faktor utama yang memengaruhi prevalensi *Tuberculosis* (TBC), seperti populasi, faktor lingkungan, akses layanan kesehatan, kepadatan penduduk, kualitas udara. Dataset mencakup informasi tentang angka dari penyakit *Tuberculosis* (TBC) di berbagai wilayah Indonesia. Selain itu, penelitian ini juga bertujuan untuk menghasilkan peta visual sebaran penyakit *Tuberculosis* (TBC) yang dapat memberikan gambaran mendalam tentang pola penyebaran penyakit tersebut. Hasil penelitian ini diharapkan dapat mendukung pengambilan keputusan berbasis data oleh pemerintah dan pemangku kepentingan lainnya dalam upaya pengendalian *Tuberculosis* (TBC) di Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan konteks di atas, masalah penelitian ini dapat dirumuskan sebagai berikut.

1. Bagaimana cara mengidentifikasi wilayah di Indonesia dengan tingkat prevalensi penyakit *Tuberculosis* (TBC) tertinggi menggunakan algoritma *K-Means Clustering*?
2. Faktor-faktor apa saja, seperti demografi, kondisi lingkungan, dan akses kesehatan, yang paling memengaruhi tingkat prevalensi penyakit *Tuberculosis* (TBC) di Indonesia berdasarkan analisis algoritma *Support Vector Regressor*?
3. Bagaimana menghasilkan peta visual yang menggambarkan sebaran prevalensi Penyakit *Tuberculosis* (TBC) di Indonesia untuk mendukung pengambilan keputusan berbasis data?

### 1.3 Batasan Masalah

Agar penelitian ini dapat dilaksanakan dengan tepat sasaran, sesuai dengan latar belakang, rumusan masalah, tujuan penelitian, serta manfaat penelitian, maka untuk batasan masalah adalah sebagai berikut.

1. Penelitian ini hanya mencakup wilayah di Indonesia yang memiliki data lengkap terkait prevalensi TBC dan faktor-faktor pendukungnya, seperti populasi, faktor lingkungan, akses layanan kesehatan, kepadatan penduduk, dan kualitas udara. Wilayah dengan data yang tidak lengkap atau hilang akan dikecualikan dari analisis.
2. Faktor-faktor yang dianalisis dibatasi pada variabel yang tersedia dalam dataset, yaitu prevalensi TBC, populasi, faktor lingkungan, akses layanan kesehatan, kepadatan penduduk, dan kualitas udara.
3. Penelitian ini hanya menggunakan algoritma *K-Means Clustering* untuk mengelompokkan wilayah berdasarkan tingkat prevalensi TBC dan algoritma *Support Vector Regressor* untuk menganalisis faktor-faktor yang memengaruhi prevalensi TBC, Algoritma lain tidak digunakan dalam penelitian ini. Analisis akan dilakukan dengan menggunakan Google Colab sebagai platform,

menggunakan python sebagai Bahasa pemrograman, dan yang terakhir menggunakan Power BI sebagai visualisasi, sedangkan untuk tools atau platform lain tidak digunakan dalam penelitian ini.

#### 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut.

1. Mengidentifikasi wilayah dengan tingkat prevalensi penyakit *Tuberculosis* (TBC) tertinggi di Indonesia menggunakan algoritma K-Means Clustering.
2. Menganalisis faktor-faktor yang paling memengaruhi prevalensi penyakit *Tuberculosis* (TBC), seperti populasi, faktor lingkungan, akses layanan kesehatan, kepadatan penduduk, dan kualitas udara, menggunakan algoritma *Support Vector Regressor*.
3. Menghasilkan peta visual sebaran TBC untuk mendukung pengambilan keputusan berbasis data.

#### 1.5 Manfaat Penelitian

Penelitian ini mungkin bermanfaat bagi beberapa pihak diantaranya.

1. Penelitian ini dapat memberikan kontribusi ilmiah dalam pengembangan metode analisis data berbasis machine learning, khususnya dalam konteks kesehatan masyarakat, dengan menggunakan algoritma K-Means Clustering dan *Support Vector Regressor* untuk memetakan prevalensi dan menganalisis faktor-faktor yang memengaruhi *Tuberculosis* (TBC).
2. Membantu meningkatkan kesadaran masyarakat terhadap faktor-faktor risiko TB, sehingga memotivasi perilaku yang mendukung pencegahan penyakit tersebut.
3. Menghasilkan peta visual sebaran *Tuberculosis* (TBC), penelitian ini memberikan alat yang mudah dipahami oleh para pemangku kepentingan,

sehingga dapat mendukung pengambilan keputusan yang lebih cepat dan akurat.

## 1.6 Kontribusi

Dalam penelitian ini, penulis dapat memberikan beberapa kontribusi terhadap beberapa pihak, diantaranya.

1. Penelitian ini berkontribusi pada pengembangan literatur ilmiah terkait penerapan algoritma *machine learning*, khususnya *K-Means Clustering* dan *Support Vector Regressor (SVR)*, dalam analisis data kesehatan masyarakat.
2. Penelitian ini menyediakan wawasan yang mendalam untuk mendesain program intervensi yang lebih terarah dan berbasis bukti.
3. Penelitian ini menunjukkan bagaimana teknologi *machine learning* dapat diterapkan dalam pengelolaan data kesehatan, sehingga dapat mendorong inovasi teknologi di bidang analisis data kesehatan lainnya.

## 1.7 Sistematika Penulisan

Penulisan membentuk sistematika penulisan yang tersusun sebanyak 5 bab, yaitu.

1. BAB I Pendahuluan

Bab ini menjelaskan pendahuluan yang memuat latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika skripsi yang telah diidentifikasi.

2. BAB II Tinjauan Pustaka

Bab ini berisi landasan teori yang relevan dengan algoritma *K-Means Clustering* dan *Support Vector Regressor (SVR)* dalam analisis pemetaan penyakit *Tuberculosis* di Indonesia. Landasan teori akan mencakup konsep dasar *machine learning*, termasuk algoritma *K-Means Clustering* dan algoritma *Support Vector Regressor (SVR)*, serta penerapannya dalam

menganalisis data penyakit *Tuberculosis* dengan menggunakan python dan Power BI.

3. BAB III Metodologi Penelitian

Bab ini berisi uraian teknik pengumpulan data, sumber data, analisis data, hasil dan Kesimpulan analisis data.

4. BAB IV Hasil dan Pembahasan

Bab ini memuat hasil dan pembahasan penelitian.

5. BAB V Kesimpulan dan Saran

Bab ini memuat Kesimpulan dan saran dari hasil penelitian.

