

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PENGEROMBOLAN PENYEBARAN PENYAKIT MENULAR DI INDONESIA TAHUN 2023 MENGGUNAKAN *K-MEANS*++ BERBASIS DEKOMPOSISI TENSOR

ANGEL MARTHA PRADINA PANGARIBUAN



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
2025**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penggerombolan Penyebaran Penyakit Menular di Indonesia Tahun 2023 Menggunakan *K-Means++* Berbasis Dekomposisi Tensor” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2025

Angel Martha Pradina Pangaribuan
G1401211105



ABSTRAK

ANGEL MARTHA PRADINA PANGARIBUAN. Penggerombolan Penyebaran Penyakit Menular di Indonesia Tahun 2023 Menggunakan *K-Means++* Berbasis Dekomposisi Tensor. Dibimbing oleh AKBAR RIZKI dan ANIK DJURAIDAH.

Penyakit menular merupakan masalah kesehatan masyarakat yang utama karena pola penularannya yang kompleks, yang dipengaruhi oleh faktor spasial, temporal, dan individu. Pendekatan satu dimensi konvensional sering kali gagal untuk menangkap dimensi-dimensi yang saling terkait ini. Penelitian ini menggunakan dekomposisi tensor dan penggerombolan *K-Means++* untuk memeriksa penyebaran penyakit menular di Indonesia. Data diperoleh dari sampel reguler BPJS Kesehatan tahun 2023 yang terdiri dari lebih dari 3,7 juta catatan kunjungan pasien ke fasilitas kesehatan tingkat pertama di 34 provinsi. Klasifikasi penyakit mengikuti standar ICD-10 yang dikeluarkan oleh Organisasi Kesehatan Dunia (WHO). Data disusun sebagai tensor tiga arah (wilayah $\times$ jenis penyakit $\times$ waktu) dan dekomposisi *Tucker* dengan rank optimal $3 \times 3 \times 3$ dipilih berdasarkan proporsi kumulatif dan galat. Hasil menunjukkan penggerombolan dengan membentuk tujuh gerombol wilayah, empat gerombol jenis penyakit, dan tiga gerombol waktu dengan karakteristik yang berbeda. Provinsi dengan kepadatan penduduk yang tinggi seperti Jawa Barat, Jawa Tengah, dan Jawa Timur berada pada gerombol wilayah yang sama dengan rata-rata beban kasus tertinggi. Penggerombolan jenis penyakit didasarkan pada rata-rata jumlah penderita dimana penyakit seperti infeksi saluran pencernaan dan tuberkulosis memiliki rata-rata kasus tertinggi dan tergolong dalam gerombol penyakit yang sama. Sementara itu, penggerombolan berdasarkan waktu menunjukkan pola penyebaran penyakit menular dipengaruhi oleh karakteristik penyakit seperti pada bulan Mei dan Oktober sebagai puncak rata-rata banyaknya penderita penyakit menular.

Kata kunci: BPJS Kesehatan, dekomposisi tensor, *k-means++*, penggerombolan, penyakit menular

ABSTRACT

ANGEL MARTHA PRADINA PANGARIBUAN. Clustering the Spread of Infectious Diseases in Indonesia in 2023 Using K-Means++ Based on Tensor Decomposition. Supervised by AKBAR RIZKI and ANIK DJURAIDAH.

Infectious diseases are a major public health issue due to their complex transmission patterns, which are influenced by spatial, temporal, and individual factors. Conventional one-dimensional approaches often fail to capture these interrelated dimensions. This study uses Tensor Decomposition and K-Means++ clustering to examine the spread of infectious diseases in Indonesia. Data was obtained from the 2023 BPJS Health regular sample, comprising over 3.7 million patient visit records to primary healthcare facilities across 34 provinces. Disease classification follows the ICD-10 standard issued by the World Health Organization (WHO). Data is organized as a three-way tensor (region $\times$ disease type $\times$ time), and Tucker decomposition with an optimal rank of $3 \times 3 \times 3$ is selected based on cumulative proportion and error. The results show clustering into seven regional clusters, four disease type clusters, and three time clusters with distinct characteristics. Provinces with high population density, such as West Java, Central Java, and East Java, are grouped in the same regional cluster with the highest average case burden. Disease type clustering was based on the average number of patients, where diseases such as gastrointestinal infections and tuberculosis had the highest average number of cases and were grouped into the same disease cluster. Meanwhile, time-based clustering showed that the spread of infectious diseases was influenced by disease characteristics, with May and October as the peak months for the average number of infectious disease patients.

Keywords: BPJS health, clustering, infectious disease, k-means++, tensor decomposition.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PENGEROMBOLAN PENYEBARAN PENYAKIT MENULAR DI INDONESIA TAHUN 2023 MENGGUNAKAN *K-MEANS*++ BERBASIS DEKOMPOSISI TENSOR

ANGEL MARTHA PRADINA PANGARIBUAN

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

IPB University

Judul Skripsi : Penggerombolan Penyebaran Penyakit Menular di Indonesia
Tahun 2023 Menggunakan *K-Means++* Berbasis Dekomposisi
Tensor

Nama : Angel Martha Pradina Pangaribuan
NIM : G1401211105

Disetujui oleh

Pembimbing 1:
Akbar Rizki S.Stat., M.Si.

Pembimbing 2:
Prof. Dr. Ir. Anik Djuraidah M.S.

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 197804112005011002



Tanggal Ujian:
21 Juli 2025

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian ini "Penggerombolan Penyebaran Penyakit Menular di Indonesia Tahun 2023 Menggunakan *K-Means++* Berbasis Dekomposisi Tensor". Penulis menyadari bahwa terselesaikannya karya ilmiah ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada seluruh pihak yang telah berperan sejak awal perjalanan akademik hingga penyusunan karya ilmiah ini selesai, antara lain sebagai berikut:

1. Alm. Bapak Posma Pangaribuan, Ibu Julesna Hutagalung, Alfhonso, Cindy, serta seluruh keluarga besar yang telah memberi restu, doa, dan dukungan.
2. Ibu Akbar Rizki S.Stat., M.Si dan Ibu Prof. Dr. Ir. Anik Djuraidah M.S selaku komisi pembimbing skripsi yang telah memberikan banyak arahan dan saran dalam penulisan karya ilmiah ini.
3. Ibu Aulia Rizki Firdawanti, S.Stat., M.Si selaku moderator kolokium, Ibu Sachnaz Desta Oktarina, S.Stat, M.Agr., Ph.D selaku moderator seminar, dan Ibu Pika Silvianti, S.Si., M.Si selaku dosen penguji yang telah memberikan kritik serta masukkan untuk perbaikan karya ilmiah ini.
4. Seluruh dosen dan staf akademik Prodi S1 Statistika dan Sains Data, SSMI yang telah menyampaikan ilmunya dan membantu selama proses perkuliahan hingga menyelesaikan studi.
5. Dhiya, Diva, Farhan, Fuadiyah, Kheni, dan Reyzha selaku teman seperjuangan penulis selama berkuliah yang senantiasa membantu, memberikan dukungan, dan menciptakan banyak kenangan berharga selama masa studi.
6. Teman-teman Statistika 58, khususnya Adiba, Arfiah, Azra, Fajrin, Farras, Fedora, Gladys, Nadila, Reza, dan Vita yang telah menemani dan mewarnai masa perkuliahan.
7. Roosalva Ria Stevany, sahabat sejak SD yang senantiasa mendukung dan memberikan motivasi hingga kini.
8. Aqilah dan Sharikha selaku teman penulis yang telah menjadi rekan baik dan penyemangat dari masa SMA hingga sekarang.
9. Teman-teman dari tim *Survey and Research* GSB 2024, Arutala Tabbena PMB IPB 60, serta teman satu bimbingan yang telah menjadi bagian penting dalam perjalanan pembelajaran dan pengalaman selama masa studi.

Bogor, Juli 2025

Angel Martha Pradina Pangaribuan

DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Dekomposisi Tensor	3
2.2 Dekomposisi <i>Tucker</i>	4
2.3 <i>Higher-Order Orthogonal Iteration</i> (HOOI)	4
2.4 Penggerombolan <i>K-Means++</i>	4
2.5 Penyakit Menular	6
III METODE	7
3.1 Data	7
3.2 Prosedur Analisis Data	8
IV HASIL DAN PEMBAHASAN	13
4.1 Praproses Data	13
4.2 Dekomposisi Tensor	15
4.3 Penggerombolan <i>K-Means++</i>	16
4.4 Penggerombolan Wilayah	17
4.5 Penggerombolan Jenis Penyakit	21
4.6 Penggerombolan Waktu	23
V SIMPULAN	25
5 Simpulan	25
DAFTAR PUSTAKA	26
LAMPIRAN	30
RIWAYAT HIDUP	39

DAFTAR TABEL

1	Peubah yang digunakan	7
2	Ragam kombinasi rank antardimensi	15
3	Daftar anggota gerombol dimensi wilayah	17
4	Statistika deskriptif gerombol wilayah	18
5	Daftar anggota gerombol dimensi jenis penyakit	21
6	Statistika deskriptif gerombol jenis penyakit	22
7	Daftar anggota gerombol waktu	23
8	Statistika deskriptif gerombol waktu	23

DAFTAR GAMBAR

1	Susunan elemen berdasarkan setiap dimensi tensor	9
2	Ilustrasi tensor tiga dimensi: wilayah $\times$ jenis penyakit $\times$ waktu	9
3	Hasil dekomposisi <i>Tucker</i>	11
4	Peta jumlah peserta penyakit menular per provinsi berdasarkan data BPJS Reguler tahun 2023 (peta tidak berskala)	14
5	Diagram batang penyakit menular tertinggi di Indonesia berdasarkan data BPJS Reguler tahun 2023	14
6	Diagram garis total penyakit menular per bulan berdasarkan data BPJS Reguler tahun 2023	15
7	Plot galat rekonstruksi untuk rank $3 \times 3 \times 3$	16
8	Grafik <i>silhouette</i> pada setiap dimensi	16
9	Peta hasil penggerombolan per provinsi	17

DAFTAR LAMPIRAN

1	Daftar penyakit menular	31
2	Contoh hasil perubahan format tanggal ke bulan	31
3	Contoh data setelah diurutkan berdasarkan provinsi, penyakit, bulan, dan nomor peserta	32
4	Contoh data duplikat berdasarkan kombinasi bulan dan jenis penyakit	32
5	Contoh hasil perhitungan total bobot	33
6	Penyakit dominan dan waktu lonjakan kasus per gerombol wilayah	33
7	Penyakit dominan dan wilayah lonjakan kasus per gerombol waktu	37