

ABSTRAK

Mutia Rizka Nohu, 2025. *Perbandingan Metode Jarak Gower, Log-Likelihood dan Tipe Data Campuran pada Algoritma K-Prototype untuk Mengelompokkan Provinsi di Indonesia.* **SKRIPSI.** Gorontalo. Program Studi Statistika. Jurusan Matematika. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Negeri Gorontalo. Pembimbing : **(1) Nurwan, S.Pd., M.Si., (2) Siti Nurmardia Abdussamad, S.Stat., M.Stat.**

Clustering adalah metode analisis untuk mengelompokkan data berdasarkan kesamaan karakteristik. Salah satu metode *clustering* non-hirarki adalah algoritma *K-Prototype* yang merupakan penggabungan algoritma *K-Means* dan *K-Modes*, digunakan untuk menangani data campuran numerik dan kategorik. Pemilihan metode jarak yang tepat sangat mempengaruhi hasil *clustering*. Penelitian ini membandingkan efektivitas jarak Gower, *Log-likelihood* dan tipe data campuran dengan algoritma *K-Prototype* untuk mengelompokkan provinsi di Indonesia berdasarkan Indeks Ketimpangan Gender (IKG). Ketimpangan gender di Indonesia masih menjadi hambatan dalam mencapai kesetaraan gender, yang berdampak pada pertumbuhan ekonomi dan sosial. Hasil penelitian ini diharapkan dapat memberikan wawasan mengenai pengelompokan provinsi berdasarkan IKG untuk mendukung kebijakan yang lebih tepat dalam mengatasi kesenjangan gender. Penelitian ini menggunakan data sekunder dari laman BPS tahun 2023, dengan populasi adalah IKG dan variabel-variabel yang menggambarkan IKG pada 34 provinsi di Indonesia tahun 2023, dengan metode sampling jenuh yang melibatkan seluruh anggota populasi sebagai sampel. Hasil penelitian menunjukkan bahwa metode jarak Gower lebih baik dari kedua metode jarak lainnya dengan nilai *Silhouette Coefficient* sebesar 0.48 dengan jumlah *cluster* optimal sebanyak tiga. *Clustering* dengan metode jarak Gower menunjukkan bahwa *cluster* 1 terdiri dari provinsi dengan IKG rendah, *cluster* 2 dengan IKG sedang sedangkan *cluster* 3 memiliki tingkat IKG tinggi. Penelitian ini menunjukkan bahwa kesenjangan gender masih cukup signifikan di berbagai provinsi, sehingga diperlukan kebijakan berbasis data yang lebih spesifik untuk mengatasi ketimpangan, terutama pada provinsi dalam *cluster* 3.

Kata Kunci: Gower, *Log-likelihood*, Tipe Data Campuran, *K-Prototype*, Ketimpangan Gender

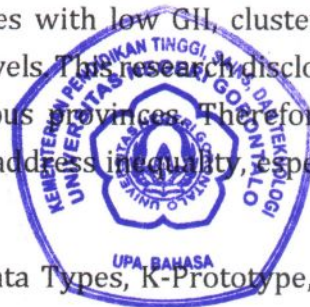
ABSTRACT

Mutia Rizka Nohu, 2025. *Comparison of Gower's Distance, Log-Likelihood, and Mixed Data Type Methods in the K-Prototype Algorithm for Grouping Provinces in Indonesia.* **Undergraduate Thesis.** Gorontalo. Study Program of Statistics. Department of Mathematics. Faculty of Mathematics and Natural Sciences. Universitas Negeri Gorontalo.

The Supervisors: **(1) Nurwan, S.Pd., M.Si., (2) Siti Nurmardia Abdussamad, S.Stat., M.Stat.**

Clustering is an analysis method for grouping data based on the similarity of characteristics. One of the non-hierarchical clustering methods is the K-Prototype algorithm, a combination of K-Means and K-Modes algorithms. This method is used to handle numerical and categorical mixed data because selecting the proper distance method significantly affects the clustering results. This study compared the effectiveness of Gower's distance, Log-likelihood, and mixed data types with the K-Prototype algorithm to classify provinces in Indonesia based on the Gender Inequality Index (GII). Gender inequality in Indonesia is still an obstacle to achieving gender equality because it has an impact on economic and social growth. The results of this study are expected to provide insight into the grouping of provinces based on GII to support more appropriate policies in overcoming gender gaps. This study used GII as the population and secondary data from the Indonesia Statistics website in 2023, with variables describing the GII in 34 provinces in Indonesia in 2023. The sample was taken using a total sampling involving all members of the population as a sample. Based on the analysis, Gower's distance method was better, with a Silhouette Coefficient value of 0.48 and three optimal numbers of clusters. Clustering with Gower's distance method showed that cluster 1 consisted of provinces with low GII, cluster 2 with moderate, and cluster 3 with the highest GII levels. This research disclosed that the gender gap was still significant in various provinces. Therefore, more specific data-driven policies are required to address inequality, especially in provinces in cluster 3.

Keywords: Gower, Log-likelihood, Mixed Data Types, K-Prototype, Gender Inequality



LEMBAR PENGESAHAN

Skripsi yang berjudul "**PERBANDINGAN METODE JARAK GOWER,
LOG-LIKELIHOOD DAN TIPE DATA CAMPURAN PADA ALGORITMA
K-PROTOTYPE UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA**"

Oleh

MUTIA RIZKA NOHU
NIM. 413421041

Program Studi Statistika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Telah dipertahankan di depan pembimbing dan dewan penguji

Hari, tanggal : Rabu, 28 Mei 2025

Waktu : 15.01-16.00 WITA

Tempat : Ruang Sidang Statistika

A. Pembimbing

1. **Nurwan, S.Pd., M.Si.**

NIP. 198105102006041002

2. **Siti Nurmardia Abdussamad, S.Stat., M.Stat.**

NIP. 199503042024212001

B. Dewan Penguji

3. **Djihad Wunguli, S.Pd., M.Si.**

NIP. 198906122019031018

4. **Armayani Aarsal, S.Si., M.Si.**

NIP. 199505142022032016

Tanda Tangan

Pembimbing I

(.....)

Pembimbing II

(.....)

Tanda Tangan

Penguji I

(.....)

Penguji II

(.....)

Mengetahui,

Dekan Fakultas Matematika dan IPA



Prof. Dr. Fitriyane Lihawa, M.Si.

NIP.196912091993032001