IMPLEMENTASI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *GAP STATISTIC* UNTUK *CLUSTERING* KANKER PAYUDARA PADA ALGORITMA *K-MEANS*

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh
RIDHA AFIFA
NIM 1911016320015

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

JANUARI 2024

IMPLEMENTASI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *GAP STATISTIC* UNTUK *CLUSTERING* KANKER PAYUDARA PADA ALGORITMA *K-MEANS*

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh
RIDHA AFIFA
NIM 1911016320015

PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

JANUARI 2024

SKRIPSI

IMPLEMENTASI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *GAP STATISTIC* UNTUK *CLUSTERING* KANKER PAYUDARA PADA ALGORITMA *K-MEANS*

Oleh:
RIDHA AFIFA
NIM 1911016320015

Telah dipertahankan di depan Dosen Penguji pada Tanggal 4 Januari 2024

Susunan Dosen Penguji:

Pembimbing I

M. Itqan Mazdadi, S.Kom, M.Kom.
NIP. 199006122019031013

Dosen Penguji I

Fatma Indriani, S.T., M.I.T., Ph.D.
NIP. 198404202008122004

Pembimbing II

Triando H. Saragih, S.Kom., M.Kom.
NIP. 199308242019031012

Dosen Penguji II

Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002

Banjarbaru, 4 Januari 2024
Koordinator Program Studi Ilmu Komputer

Irwan Budiman, S.T., M.Kom.
NIP. 19770325 2008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diberikan orang lain, kecuali yang secara tertulis diacu di dalam naskah ini dan disebutkan dalam Daftar Pustaka

Banjarbaru, 4 Januari 2024

RIDHA AFIFA
NIM. 1911016320015

ABSTRAK

IMPLEMENTASI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *GAP STATISTIC* UNTUK *CLUSTERING* KANKER PAYUDARA PADA ALGORITMA *K-MEANS*

(Oleh: Ridha Afifa; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Triando Hamonangan Saragih, S.Kom., M.Kom.; 2024; 79 halaman)

Kanker payudara menjadi salah satu penyebab kematian paling umum di dunia. Kanker payudara dapat dideteksi menggunakan *data mining*, di mana data diekstraksi menjadi informasi yang berguna. *Clustering* kanker payudara dilakukan untuk membantu pihak medis dalam mengelompokkan karakteristik setiap jenis kanker. Namun, pada data kanker payudara terdapat multikolinieritas data sehingga dapat mempengaruhi hasil *clustering*. Untuk menangani masalah tersebut ditangani menggunakan reduksi dimensi *Principal Component Analysis* (PCA). Metode *Principal Component Analysis* dapat mengatasi masalah multikolinieritas data dan meningkatkan efisiensi komputasi. Selain itu metode *K-Means* juga memiliki kelemahan dalam menentukan jumlah kluster yang optimal, sehingga digunakan metode *Gap Statistic* untuk mencari nilai K optimal yang cocok digunakan pada data kanker payudara. Dalam penelitian ini, dilakukan perbandingan hasil evaluasi dari model *clustering K-Means*, gabungan model *clustering PCA-KMeans* dan gabungan model *clustering PCA-GapStatistic-KMeans*. Dari penelitian ini, didapatkan hasil evaluasi pada model *clustering K-Means* dengan reduksi dimensi PCA dan K optimal *Gap Statistic* lebih baik dibandingkan model *K-Means* tanpa reduksi dimensi. Dengan jumlah kluster yang dihasilkan oleh *Gap Statistic* sebanyak 2 kluster dan hasil evaluasi yang diperoleh sebesar 1.195513.

Kata kunci: Kanker payudara, *clustering, K-Means, Principal Component Analysis, Gap Statistic*

ABSTRACT

IMPLEMENTATION OF PRINCIPAL COMPONENT ANALYSIS (PCA) AND GAP STATISTIC FOR BREAST CANCER CLUSTERING IN THE K-MEANS ALGORITHM *(By: Ridha Afifa; Supervisor: Muhammad Itqan Mazdadi, S.Kom., M.Kom. and Triando Hamonangan Saragih, S.Kom., M.Kom.; 2024; 79 pages)*

Breast cancer is one of the most common causes of death worldwide. Data mining can be utilized to detect breast cancer, where information is extracted from data to provide valuable insights. Clustering of breast cancer is conducted to assist medical professionals in grouping the characteristics of each cancer type. However, multicollinearity in breast cancer data can impact clustering results. To address this issue, dimensionality reduction through Principal Component Analysis (PCA) is employed. PCA can effectively handle multicollinearity issues and enhance computational efficiency. Additionally, the K-Means method has limitations in determining the optimal number of clusters. Therefore, the Gap Statistic method is employed to find the optimal K value suitable for breast cancer data. This study compares the evaluation results of the K-Means clustering model, the combined PCA-KMeans clustering model, and the combined PCA-GapStatistic-KMeans clustering model. The findings indicate that the evaluation results for the K-Means model with PCA dimensionality reduction and optimal Gap Statistic K are superior to the K-Means model without dimensionality reduction. The Gap Statistic suggests 2 clusters as the optimal number, with an evaluation result of 1.195513.

Keywords: Breast cancer, clustering, K-Means, Principal Component Analysis, Gap Statistic

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul "Implementasi *Principal Component Analysis* (PCA) Dan *Gap Statistic* Untuk *Clustering* Kanker Payudara Pada Algoritma *K-Means*" untuk memenuhi syarat dalam menyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiyamah.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga terutama orang tua yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Muhammad Itqan Mazdadi S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Teman-teman keluarga Ilmu Komputer angkatan 2019, terima kasih untuk canda, tawa, perjuangan yang sudah dilewati bersama, untuk semua kenangan manis yang telah terukir selama ini. Senang bisa menjadi salah satu bagian dari kehidupan kalian.
6. Ucapan terimakasih kepada Rabiatun Nisa, Zaina Fadia, Hartati, Nurmolika, dan Maimunah yang telah membersamai melewati lika liku perkuliahan serta memberikan dukungan dalam proses pengerjaan skripsi.

7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka saran dan kritik dari semua pihak sangat diharapkan demi penyempurnaan pada penelitian selanjutnya. Diharapkan skripsi ini dapat bermanfaat bagi semua pihak dan dapat menambah pengetahuan kita semua.

Banjarbaru, 4 Januari 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN