

**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* (PCA) PADA KLASIFIKASI POLA *BEAT* EKG
PENYAKIT ARITMIA**

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**

Oleh :

Mira Sulistiana

NIM. 08011381924082



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
TAHUN 2023**

LEMBAR PENGESAHAN

**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* (PCA) PADA KLASIFIKASI POLA *BEAT* EKG
PENYAKIT ARITMIA**

SKRIPSI

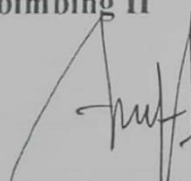
**Sebagai Salah Satu Syarat untuk
Memperoleh Gelar Sarjana Matematika**

Oleh

**MIRA SULISTIANA
NIM. 08011381924082**

Indralaya, 28 September 2023


Pembimbing II


Dr. Anita Desiani, S.Si., M.Kom.
NIP. 197712112003122002

Pembimbing I


Drs. Sugandi Yahdin, M.M.
NIP. 19580727 198603 1003

**Mengetahui,
Ketua Jurusan Matematika**


Dr. Dian Cahyawati, S.Si., M.Si.
NIP. 197303212000122001

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Mira Sulistiana
NIM : 08011381924082
Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan sarjana satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 28 September 2023



Penulis

LEMBAR PENGESAHAN

Kupersembahkan skripsi ini untuk :

Yang Mahakuasa Allah Subhanahu Wa Ta'ala,

Kedua orang tuaku tercinta,

Adik laki-lakiku tersayang,

Seluruh guru dan dosenku,

Sahabat-sahabatku,

Dan almamaterku.

Moto

“we’re limited, but Allah unlimited”

KATA PENGANTAR

Puji syukur atas kehadiran Allah Subhanahu Wa Ta'ala yang telah memberikan kemudahan dan kelancaran, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Reduksi Dimensi dengan *Principal Component Analysis* (PCA) pada Klasifikasi Pola *Beat* EKG Penyakit Aritmia” sebagai salah satu syarat memperoleh gelar sarjana sains bidang studi Matematika di Fakultas MIPA Universitas Sriwijaya. Penulis menyadari bahwa setiap proses penyusunan skripsi ini merupakan proses pembelajaran yang berharga serta tidak luput dari kekurangan dan keterbatasan. Dengan segala hormat dan kerendahan hati, penulis mengucapkan terima kasih dan penghargaan kepada Kedua orangtuaku tersayang, Bapak Samsul dan Ibu Kartini yang selalu mendoakan dan mendukung Penulis hingga saat ini dan nanti. Penulis mengucapkan terimakasih dan penghargaan kepada :

1. Bapak Prof. Hermansyah, S.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya. Bapak Drs. Sugandi Yahdin, M.M. selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya dan pembimbing utama yang telah memberikan arahan dan motivasi kepada penulis selama proses perkuliahan. Ibu Dr. Dian Cahyawati Sukanda, M.Si. selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang telah mengarahkan urusan akademik kepada penulis.

2. Ibu Dr. Anita Desani, S.Si., M.Kom. selaku dosen pembimbing kedua, Ibu Dr. Yuli Andriani, S.Si., M.Si. selaku penguji pertama dan Ibu Oki Dwipurwani, S.Si., M.Si. selaku penguji kedua yang telah bersedia memberikan waktu, tenaga, pikiran, nasehat, dan motivasi untuk memberikan bimbingan dan pengarahan selama proses perkuliahan, berkompetisi, dan yang utama selama proses penyusunan skripsi.
3. Bapak Drs. Robinson Sitepu, M.Si dan Ibu Des Alwine Zayanti, S.Si., M.Si. selaku tim pelaksana Seminar Proposal, Seminar Hasil, dan Sidang Tugas Akhir Penulis. Bapak Drs. Ali Amran, M.T. selaku dosen pembimbing akademik yang telah membimbing dan mengarahkan urusan akademik penulis.
4. Seluruh Dosen di Jurusan Matematika FMIPA yang telah memberikan ilmu, nasihat, motivasi, dan bimbingannya selama proses perkuliahan.
5. Bapak Irwansyah selaku pengurus administrasi dan ibu Hamidah selaku pengurus tata usaha Jurusan Matematika FMIPA Universitas Sriwijaya yang telah membantu penulis selama proses perkuliahan.
6. Adik laki-lakiku tersayang, Alwi Hidayat, yang selalu memberikan semangat kepada penulis, beserta keluarga besar yang selalu memberikan doa dan dorongan kepada penulis.
7. Seluruh sahabat seperjuangan, Keluarga Matematika 2019, Tim PHP2D 2020, BPH Himastik Gelora Karya, BPH Himastik Akselerasi yang telah kebersamai selama masa perkuliahan.

**DIMENSIONAL REDUCTION WITH PRINCIPAL COMPONENT
ANALYSIS (PCA) IN THE CLASSIFICATION OF ECG BEAT
PATTERNS IN ARRHYTHMIA DISEASES**

Oleh

MIRA SULISTIANA

NIM. 08011381924082

Dimensional reduction is a technique for reducing the number of features in a dataset. Principal Component Analysis (PCA) is a dimensional reduction method that can reduce the features in a dataset without eliminating important information in it. In this study, the aim is to obtain the results of dimension reduction as well as to know the performance results before and after data dimension reduction. In this study, PCA was used to reduce the dimensions of ECG beat data for arrhythmias and to speed up the classification process. The method used in this study is the PCA method to reduce large data dimensions into smaller dimension datasets so as to simplify the classification process. Measurement of performance results using the Naïve Bayes algorithm, KNN, and Decision Tree. There is an increase in performance results after the data dimensions are reduced in the Naïve Bayes algorithm, there is an increase in precision of 1%, recall of 3%, and F1-Score of 4%. In the KNN classification there is an increase in precision, recall, and F1-Score each of 0.5%. In the Decision Tree classification there is an increase in precision, recall, and F1-Score performance of 1% each. Based on the results obtained, it can be concluded that the use of PCA can improve the performance of the classification method on the ECG beat dataset for arrhythmias.

Keywords: Dimensional Reduction, PCA, ECG, Arrhythmia, Classification

**REDUKSI DIMENSI DENGAN *PRINCIPAL COMPONENT ANALYSIS* (PCA) PADA KLASIFIKASI POLA *BEAT* EKG
PENYAKIT ARITMIA**

Oleh

MIRA SULISTIANA

NIM. 08011381924082

ABSTRAK

Reduksi dimensi adalah teknik untuk mengurangi jumlah fitur pada suatu *dataset*. Principal Component Analysis (PCA) adalah metode reduksi dimensi yang dapat mengurangi fitur dalam *dataset* tanpa menghilangkan informasi penting didalamnya. Pada penelitian ini, bertujuan untuk mendapatkan hasil reduksi dimensi serta, mengetahui hasil kinerja sebelum dan sesudah reduksi dimensi data. Pada penelitian ini, PCA digunakan untuk mereduksi dimensi data beat EKG penyakit aritmia serta, mempercepat proses klasifikasi. Metode yang digunakan pada penelitian ini adalah metode PCA untuk mereduksi dimensi data yang besar menjadi *dataset* berdimensi lebih kecil sehingga dapat mempermudah proses klasifikasi. Pengukuran hasil kinerja menggunakan algoritma Naïve Bayes, KNN, dan Decision Tree. Terdapat kenaikan hasil kinerja setelah data direduksi dimensi pada algoritma Naïve Bayes terdapat kenaikan presisi sebesar 1%, recall sebesar 3%, dan F1-Score sebesar 4%. Pada klasifikasi KNN terdapat kenaikan presisi, recall, dan F1-Score masing-masing sebesar 0,5%. Pada klasifikasi Decision Tree terdapat kenaikan kinerja presisi, recall, dan F1-Score masing-masing sebesar 1%. Berdasarkan hasil yang diperoleh dapat disimpulkan penggunaan PCA dapat meningkatkan kinerja pada metode klasifikasi pada *dataset beat* EKG penyakit aritmia.

Kata Kunci : Reduksi Dimensi, PCA, EKG, Aritmia, Klasifikasi

DAFTAR ISI

HALAMAN JUDUL	i
KATA PENGANTAR	v
ABSTARK	viii
DAFTAR ISI	ix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
BAB II TINJAUN PUSTAKA.....	5
2.1 Reduksi Dimensi	5
2.2 Matriks	5
2.3 Pricipal Component Analysis (PCA)	6
2.4 Confussion Matrix.....	13
BAB III METODOLOGI PENELITIAN	17
3.1 Tempat.....	17
3.2 Waktu	17
3.3 Alat.....	17
3.4 Metode Penelitian.....	17
BAB IV HASIL DAN PENELITIAN	21
4.1 Deskripsi Data.....	21
4.2 Reduksi Dimensi	22
4.3 Perhitungan Manual PCA	22
4.4 Pengujian Kinerja Algoritma	43
4.5 Analisis dan Interpretasi Hasil	50
BAB V KESIMPULAN DAN SARAN	54
5.1 Kesimpulan	54
5.2 Saran.....	54
DAFTAR PUSTAKA	55

DAFTAR TABEL

Tabel 4.1 Data Pola Beat EKG Penyakit Aritmia	21
Tabel 4.2 Nilai Proporsi Kumulatif.....	37
Tabel 4.3 Data yang Telah Direduksi Dimensi	38
Tabel 4.4 Nilai Proporsi Kumulatif.....	42
Tabel 4.5 Dataset Pola Beat EKG Penyakit Aritmia	43
Tabel 4.6 Confussion Matrix Algoritma Naive Bayes Sebelum Reduksi Dimensi	44
Tabel 4.7 Confussion Matrix Algoritma Naive Bayes Sesudah Reduksi Dimensi	47
Tabel 4.8 Confussion Matrix Algoritma KNN Sebelum Reduksi Dimensi.....	48
Tabel 4.9 Confussion Matrix Algoritma KNN Sesudah Reduksi Dimensi	48
Tabel 4.10 Confussion Matrix Algoritma Decision Tree Sebelum Reduksi Dimens	49
Tabel 4. 11 Confussion Matrix Algoritma Decision tree Sesudah Reduksi Dimensi	50
Tabel 4.12 Perbandingan Hail Kinerja Kombinasi Algoritma.....	50
Tabel 4.13 Rata-rata Hasil Akurasi, Presisi, Recall, dan F1-Score.....	51
Tabel 4.14 Perbandingan Hasil Penelitian Terdahulu	53

DAFTAR GAMBAR

Gambar 2.1 Matriks Persegi.....	5
Gambar 2.2 Matriks Kolom	6
Gambar 2.3 Matriks Terapam	6
Gambar 4.1 Grafik Kenaikan Rata-rata Hasil Klasifikasi.....	52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining digunakan untuk mengumpulkan dan mengolah data, sehingga dapat diambil informasi yang terdapat dalam data tersebut. Data yang digunakan pada data mining menggunakan data yang berjumlah besar (Osman, 2019). Salah satu *dataset* yang memiliki dimensi data besar adalah *dataset* pola *beat* EKG penyakit aritmia yang berasal dari MIT-BIH *Arrhythmia Database* pada laman www.physionet.org, data tersebut diperoleh dari 47 pasien aritmia dengan rekaman EKG selama 30 menit dan didapatkan 650.000 titik sinyal.

Data yang besar memerlukan memori yang lebih besar dan variabel yang lebih banyak, selain itu data yang terlalu besar dapat mempengaruhi penggunaan memori komputer. Keterbatasan memori akan mempengaruhi kinerja klasifikasi menjadi tidak optimal sehingga, semakin banyak variabel input akan mengakibatkan *overfitting* pada data. Cara untuk mengatasi dimensi data yang terlalu besar, maka dimensi data harus diperkecil agar proses klasifikasi dapat berjalan dengan baik. Salah satu cara untuk mengatasi *overfitting* dengan mereduksi dimensi data.

Reduksi dimensi adalah teknik untuk mengurangi multikolinieritas pada variabel input (Maxwell *et al.*, 2018). Reduksi dimensi bekerja dengan mengurangi jumlah variabel pada *dataset* tanpa menghilangkan informasi penting dalam *dataset* tersebut. Salah satu metode untuk mereduksi dimensi adalah *Principal Component Analysis* (PCA) (Anowar *et al.*, 2021). PCA memiliki kelebihan mampu mewakili

data secara proporsional berdasarkan nilai eigen. Semakin besar nilai eigen maka semakin besar keragaman data yang diwakili oleh suatu variabel terbentuk.

Beberapa penelitian yang menerapkan PCA diantaranya Azmi dan Hermawan, (2022) pada penelitiannya yang membahas keterkaitan PCA dengan klasifikasi kualitas air menggunakan algoritma *K-Nearest Neighbor* (KNN) menghasilkan presisi sebesar 90,0% dan *recall* sebesar 91,0%, namun pada penelitiannya hanya menggunakan 21 fitur. Pada penelitian lainnya, Ardiansyah *et al.*, (2022) melakukan reduksi dimensi dengan PCA pada *dataset* KDD menggunakan algoritma SVM dan menghasilkan presisi sebesar 52% namun hasil dari klasifikasi menggunakan algoritma SVM belum baik. Pada penelitian Rahmi *et al.*, (2020) mengkombinasikan PCA dan *Bayesian network* untuk mengklasifikasi faktor-faktor yang mempengaruhi ketidakhadiran karyawan menghasilkan akurasi yang optimal sebesar 89,7% namun, penelitian ini hanya menggunakan 19 fitur.

Pada penelitian ini akan menerapkan metode PCA pada *dataset beat* EKG penyakit aritmia dari laman www.physionet.org yang memiliki 253 atribut terdiri dari 252 fitur dan 1 fitur bebas. Label yang digunakan pada penelitian ini yakni, label normal dan tidak normal. Tujuannya adalah untuk mendapatkan *dataset* dengan dimensi data yang lebih kecil akan tetapi dapat mendapatkan hasil klasifikasi yang baik. Pengujian hasil klasifikasi pada *dataset* yang dihasilkan dari penerapan PCA akan menggunakan algoritma klasifikasi Naïve Bayes, KNN, dan *Decision Tree* untuk mengukur kinerja pada *dataset* baru yang dihasilkan berdasarkan nilai akurasi, presisi, *recall*, dan *F1-Score*.

1.2 Perumusan Masalah

Bagaimana menerapkan metode PCA untuk reduksi dimensi pada *dataset beat* EKG dengan 253 atribut yang terdiri atas 252 fitur bebas dan 1 label selain itu, dan melihat pengaruh PCA pada klasifikasi berdasarkan nilai akurasi, presisi, *recall*, dan *F1-Score*.

1.3 Batasan Masalah

Batasan masalah dari penelitian ini sebagai berikut:

1. Metode klasifikasi Naïve Bayes, KNN, dan *Decision Tree* dibatasi hanya untuk membandingkan hasil PCA.
2. Pengujian kinerja penilaian akhir diukur berdasarkan nilai akurasi, presisi, *recall*, dan *F1-Score*.

1.4 Tujuan

Tujuan dari penelitian ini adalah mendapatkan hasil kinerja dari penerapan kombinasi reduksi dimensi dengan PCA dan mengetahui pengaruh pada hasil klasifikasi.

1.5 Manfaat

Manfaat dari hasil penelitian ini sebagai berikut:

1. Mendapatkan *dataset* baru dengan dimensi yang lebih kecil untuk digunakan pada proses klasifikasi pola *beat* EKG penyakit aritmia.

2. Memperoleh referensi metode-metode yang dapat digunakan untuk memanfaatkan PCA pada klasifikasi *dataset* pola *beat* EKG penyakit aritmia.

DAFTAR PUSTAKA

- Ahmad Harmain, Paiman, P., Kurniawan, H., Kusriani, K., & Dina Maulina. (2022). Normalisasi data untuk efisiensi K-Means pada pengelompokan wilayah berpotensi kebakaran hutan dan lahan berdasarkan sebaran titik panas. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, 2(2), 83–89. <https://doi.org/10.46764/teknimedia.v2i2.49>
- Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, 40, 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>
- Ardiansyah, F., Hamdan, F., Sugiyanto, S., & Wahyu Siadi, I. (2022). Klasifikasi customer relationship management menggunakan dataset KDD cup 2009 dengan Teknik Reduksi Dimensi. *Komputika : Jurnal Sistem Komputer*, 11(2), 193–202. <https://doi.org/10.34010/komputika.v11i2.6498>
- Azmi, B. N., & Hermawan, A. (2022). Analisis pengaruh PCA pada klasifikasi kualitas air menggunakan algoritma K-Nearest Neighbor dan Logistic Regression. (*Jurnal Sistem Dan Informasi*, 7(2), 94–103. <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/8190%0Ahttp://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/download/8190/4143>
- Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis performa metode K-Nearest Neighbor untuk identifikasi jenis kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269–274. <https://doi.org/10.33096/ilkom.v11i3.489.269-274>
- Chen, R., Wang, D., & Xialu, L. (2022). Factor models for high-dimensional tensor Time Series. *Journal of the American Statistical Association*, 117(537), 94–116. <https://doi.org/10.1080/01621459.2021.1912757>
- Febrina, R. (2018). Faktor-faktor yang mempengaruhi jumlah kelahiran di provinsi sumatera barat dengan menggunakan Analisis Faktor. *University of Padang*, II.
- Fourure, D., Javaid, M. U., Posocco, N., & Tihon, S. (2021). Anomaly detection: how to artificially increase your F1-Score with a Biased Evaluation Protocol. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12978 LNAI*. Springer International Publishing. https://doi.org/10.1007/978-3-030-86514-6_1
- Fredicia, F., Buono, A., & Giri, E. P. (2017). Pengembangan model pengenalan wajah manusia dengan Teknik Reduksi Dimensi Bi2DPCA dan Support Vector Machine sebagai Classifier. *Jurnal ULTIMATICS*, 8(1), 11–15.

<https://doi.org/10.31937/ti.v8i1.497>

- Kim, Y. K., Lee, M., Song, H. S., & Lee, S. W. (2022). Automatic cardiac arrhythmia classification using Residual Network Combined with Long Short-Term Memory. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–17. <https://doi.org/10.1109/TIM.2022.3181276>
- Krallinger, M., Leitner, F., Rabal, O., Vazquez, M., Oyarzabal, J., & Valencia, A. (2015). CHEMDNER: The drugs and chemical names extraction challenge. *Journal of Cheminformatics*, 7(Suppl 1), 1–11. <https://doi.org/10.1186/1758-2946-7-S1-S1>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform manifold approximation and projection for Dimension Reduction. *Cornell University*, 3. <http://arxiv.org/abs/1802.03426>
- Nasution, M. Z. (2020). Face Recognition based feature extraction using Principal Component Analysis (PCA). *Journal of Informatics and Telecommunication Engineering*, 3(2), 182–191. <https://doi.org/10.31289/jite.v3i2.3132>
- Nuraliza, H., Nurul Pratiwi, O., & Hamami, F. (2022). Analisis sentimen IMDb film review dataset menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Manajemen*, 7(1), 1–17.
- Osman, A. S. (2019). Data mining techniques: Review. *International Journal of Data Science Research*, 2(1), 1–4.
- Pujianto, R., Adiwijaya, & Rahmawati, A. A. (2019). Analisis ekstraksi fitur Principle Component Analysis pada klasifikasi microarray data menggunakan Classification And Regression Trees. *EProceedings ...*, 6(1), 2368–2379.
- R Ganesh Kumar. (2014). Performance analysis of soft computing techniques for classifying cardiac arrhythmia. *Indian Journal of Computer Science and Engineering*, 4(6), 459–465.
- Rahmi, S., Sirait, P., & Panjaitan, E. S. (2020). Optimasi Klasifikasi Bayesian Network melalui reduksi attribute menggunakan metode Principal Component Analysis. *Jurnal Media Informatika Budidarma*, 4(4), 955–962. <https://doi.org/10.30865/mib.v4i4.2370>
- Sanjaya, Y. C., Gunawan, A. A. S., & Irwansyah, E. (2020). Sematic Segmentation for Aerial Images: A literature Review. *Engineering, Mathematic and Computer Science (EMACS) Journal*, 2(3), 133-139. <https://doi.org/10.21512->

/emacsjournal.v2i3.6737.

- Xie, S., Sun, C., Huang, J., Tu, Z., & Murphy, K. (2018). Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11219 LNCS(1), 318–335. https://doi.org/10.1007/978-3-030-01267-0_19
- Zhang, S., Zhang, X., Yu, W., Lin, Z., & Chen, D. (2020). Infection biomarkers in assisting the judgement of blood stream infection and patient prognosis: a retrospective study incorporating principal components analysis. *Annals of Translational Medicine*, 8(23), 1581–1581. <https://doi.org/10.21037/atm-20-3425>
- Zulfa, I., & Winarko, E. (2017). Sentimen analisis tweet berbahasa indonesia dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187. <https://doi.org/10.22146/ijccs.24716>