



PERBANDINGAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) PADA *RANDOM FOREST* UNTUK KLASIFIKASI POSISI PEMAIN SEPAK BOLA LIGA EROPA MUSIM 2022-2023

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

FADHILA NISRINA SAUMY

NIM 1911016120009

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

OKTOBER 2024



PERBANDINGAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) PADA *RANDOM FOREST* UNTUK KLASIFIKASI POSISI PEMAIN SEPAK BOLA LIGA EROPA MUSIM 2022-2023

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

FADHILA NISRINA SAUMY

NIM 1911016120009

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

OKTOBER 2024

SKRIPSI

PERBANDINGAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) PADA *RANDOM FOREST* UNTUK KLASIFIKASI POSISI PEMAIN SEPAK BOLA LIGA EROPA MUSIM 2022-2023

Oleh:

FADHILA NISRINA SAUMY

NIM. 1911016120009

Telah dipertahankan di depan Dosen Penguji pada tanggal 17 Oktober 2024.

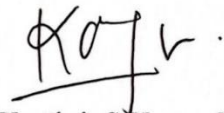
Susunan Dosen Penguji:

Pembimbing I



Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

Pembimbing II



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

Dosen Penguji II



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

Samarbaru, 30 Oktober 2024

Koordinator Program Studi Ilmu Komputer



Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

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 diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Oktober 2024

Yang Menyatakan,



Fadhila Nisrina Saumy

NIM.1911016120009

ABSTRAK

PERBANDINGAN REDUKSI DIMENSI *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *LINEAR DISCRIMINANT ANALYSIS* (LDA) PADA *RANDOM FOREST* UNTUK KLASIFIKASI POSISI PEMAIN SEPAK BOLA LIGA EROPA MUSIM 2022-2023

(Oleh : Fadhila Nisrina Saumy; Pembimbing: Irwan Budiman, S.T., M.Kom. dan Fatma Indriani, S.T., M.I.T., Ph.D; 2024; 70 halaman)

Sepak bola merupakan salah satu cabang olahraga yang populer di dunia. Olahraga ini memiliki alur permainan yang dinamis serta taktik yang cukup kompleks, oleh karena itu dibutuhkan pemahaman yang baik mengenai peran tiap posisi pemain dalam sebuah tim. Dalam beberapa tahun terakhir, analisis permainan dibantu oleh teknologi yang ada. Salah satu teknologi yang dapat digunakan dalam analisis permainan adalah penggunaan algoritma klasifikasi untuk mengelompokkan pemain berdasarkan posisi yang cocok untuk mereka. Data yang diklasifikasi bisa berupa data statistik pemain maupun data visual permainan. Fitur pada data statistik bisa tersedia secara banyak sehingga untuk mempermudah proses pengelompokan perlu dilakukan reduksi dimensi. Pada penelitian ini dilakukan klasifikasi menggunakan algoritma Random Forest dengan reduksi dimensi menggunakan algoritma Principal Component Analysis (PCA) dan Linear Discriminant Analysis (LDA) di mana hasil klasifikasi setelah mengalami reduksi dimensi akan dibandingkan. Hasil yang didapatkan dari model klasifikasi tanpa reduksi dimensi adalah sebesar 0.913 untuk akurasi dan 0.885 untuk AUC, hasil dari model klasifikasi dengan reduksi dimensi PCA adalah sebesar 0.882 untuk akurasi dan 0.807 untuk AUC, sedangkan hasil dari model klasifikasi dengan reduksi dimensi LDA adalah sebesar 0.937 untuk akurasi dan 0.95 untuk AUC.

Kata kunci: Klasifikasi, Random Forest, Reduksi Dimensi, PCA, LDA

ABSTRACT

COMPARISON OF DIMENSION REDUCTION OF PRINCIPAL COMPONENT ANALYSIS (PCA) AND LINEAR DISCRIMINANT ANALYSIS (LDA) ON RANDOM FOREST FOR CLASSIFICATION OF FOOTBALL PLAYERS POSITIONS IN EUROPEAN LEAGUE SEASON 2022-2023 (By : Fadhila Nisrina Saumy; Supervisor: Irwan Budiman, S.T., M.Kom. and Fatma Indriani, S.T., M.I.T., Ph.D; 2024; 70 pages)

Football is one of the most popular sports in the world. The sport has a dynamic game flow and quite complex tactics. Therefore, a good understanding of each player's position in a team is required. In recent years, game analysis has been assisted by existing technology. One technology that can be used in-game analysis is classification algorithms to group players based on the position that suits them. The classified data can be in the form of player statistics or visual game data. Features in statistical data can be available in large quantities, so dimensionality reduction is needed to facilitate the grouping process. In this study, classification was carried out using the Random Forest algorithm with dimensionality reduction using the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) algorithms, where the classification results after experiencing dimensionality reduction will be compared. The results obtained from the classification model without dimension reduction are 0.913 for accuracy and 0.885 for AUC. The results of the classification model with PCA dimension reduction are 0.882 for accuracy and 0.807 for AUC, while the results of the classification model with LDA dimension reduction are 0.937 for accuracy and 0.95 for AUC.

Keywords: *Classification, Random Forest, Dimension Reduction, PCA, LDA*

PRAKATA

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “PERBANDINGAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS (PCA) DAN LINEAR DISCRIMINANT ANALYSIS (LDA) PADA RANDOM FOREST UNTUK KLASIFIKASI POSISI PEMAIN SEPAK BOLA LIGA EROPA MUSIM 2022-2023” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

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

1. Ibu, Ayah, Akmal, dan Ayu Dini yang selalu memberikan bantuan, semangat, doa, dan dukungan dalam proses penyelesaian skripsi ini
2. Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Fatma Indriani, S.T., M.I.T., Ph.D selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman Ilmu Komputer angkatan 2019 yang selalu memberikan dukungan dan doa dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada Caca, Zia, Mira, Ajwa, Dina, Tina, Alya, Meftah, Yaya, Falah, Deni, Diva, Bella, dan Oca yang telah senantiasa menjadi pengingat dalam proses mengerjakan skripsi.

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

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Oktober 2024



Fadhila Nisrina Saumy

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN.....	i
HALAMAN PERNYATAAN	ii
ABSTRAK	iii
<i>ABSTRACT</i>	iv
PRAKATA	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN	
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan.....	4
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	
2.1. Literatur Terdahulu.....	6
2.2. Data Mining.....	8
2.3. Klasifikasi.....	8
2.4. Random Forest.....	8
2.5. Principal Component Analysis (PCA).....	9
2.6. Linear Discriminant Analysis (LDA).....	11
2.7. Missing Value.....	12
2.8. One-hot Encoding.....	12
2.9. Standard Scaler Normalization	13
2.10. Split Data.....	13

2.11. Confusion Matrix.....	14
2.12. Area Under The Curve (AUC)	15
2.13. Keaslian Penelitian	15
BAB III METODE PENELITIAN	
3.1. Bahan Penelitian	18
3.2. Alat Penelitian	18
3.3. Variabel Penelitian.....	18
3.4. Prosedur Penelitian	18
BAB IV HASIL DAN PEMBAHASAN	
4.1. Hasil.....	21
4.1.1. Inisialisasi Dataset	21
4.1.2. Preprocessing Dataset	33
4.1.3. Split Data	48
4.1.4. Klasifikasi	48
4.1.5. Evaluasi.....	55
4.2. Pembahasan	65
BAB V PENUTUP	
5.1. Kesimpulan.....	68
5.2. Saran	68
DAFTAR PUSTAKA	69
LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Kategori Nilai AUC	15
Tabel 2. Keaslian Penelitian.....	16
Tabel 3. Perancangan Penelitian	17
Tabel 4. Contoh Dataset Statistik Pemain Sepak Bola	21
Tabel 5. Rincian Dataset	21
Tabel 6. Deskripsi Fitur Dataset.....	22
Tabel 7. Contoh Baris dengan Missing Value	33
Tabel 8. Rincian Fitur dengan Missing Value	34
Tabel 9. Dataset Sebelum Imputasi Nilai Mean.....	35
Tabel 10. Dataset Sesudah Imputasi Nilai Mean	35
Tabel 11. Fitur yang Melalui Proses One-hot Encoding.....	36
Tabel 12. Dataset Sebelum One-hot Encoding	42
Tabel 13. Dataset Setelah One-hot Encoding	42
Tabel 14. Dataset Sebelum Normalisasi	43
Tabel 15. Dataset Setelah Normalisasi.....	43
Tabel 16. Hasil Perhitungan Matriks Kovarians	44
Tabel 17. Nilai Eigen dan Variance Cumulative	44
Tabel 18. Hasil Reduksi Dimensi PCA.....	45
Tabel 19. Hasil Persebaran Dalam Kelas (S_w)	46
Tabel 20. Hasil Persebaran Antarkelas (S_b).....	47
Tabel 21. Hasil Reduksi Dimensi LDA	47
Tabel 22. Data Acak Random Forest Manual.....	49
Tabel 23. Proporsi Kelas dengan Kriteria Perhitungan Pertama Fitur AerWon	50
Tabel 24. Proporsi Kelas dengan Kriteria Perhitungan Pertama Fitur AerLost.....	51
Tabel 25. Data Internal Node Sebelah Kiri.....	52
Tabel 26. Data Internal Node Sebelah Kanan	52
Tabel 27. Parameter Random Forest pada Scikit-learn.....	54
Tabel 28. Rincian Confusion Matrix.....	57
Tabel 29. Rincian Hasil Hitung Confusion Matrix	61

Tabel 30. Rincian Nilai AUC.....	65
Tabel 31. Jumlah Fitur, Nilai Akurasi, dan AUC Model Penelitian	67

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Perbedaan Representasi Data Numerikal dan Data One-hot Encoded dari Data Kategorikal	13
Gambar 2. Alur Penelitian Menggunakan Algoritma Random Forest dengan Reduksi Dimensi PCA dan LDA	19
Gambar 3. Jumlah Data pada Tiap Kelas Target	33
Gambar 4. Jumlah Data pada Data Latih dan Data Uji	48
Gambar 5. Persebaran Data Per Kelas	48
Gambar 6. Rata-rata Pasangan Data Tiap Fitur	50
Gambar 7. Gini Split Tiap Fitur	51
Gambar 8. Root Node Pohon Keputusan	52
Gambar 9. Gini Split AerWon Tiap Internal Node	53
Gambar 10. Pohon Keputusan Manual	53
Gambar 11. Confusion Matrix Random Forest	56
Gambar 12. Confusion Matrix Random Forest dengan PCA	56
Gambar 13. Confusion Matrix Random Forest dengan LDA	57
Gambar 14. Grafik Precision	62
Gambar 15. Grafik Recall	62
Gambar 16. Grafik F1	63
Gambar 17. Grafik Akurasi	63
Gambar 18. Kurva AUC Random Forest Tanpa Reduksi Dimensi	64
Gambar 19. Kurva AUC Random Forest dengan PCA	64
Gambar 20. Kurva AUC Random Forest dengan LDA	64
Gambar 21. Grafik AUC	65
Gambar 22. Chart Nilai Akurasi dan AUC	67

DAFTAR LAMPIRAN

Lampiran

- Lampiran 1. Source Code Import Library
- Lampiran 2. Source Code Import Dataset
- Lampiran 3. Source Code Menghapus Beberapa Baris dan Kolom
- Lampiran 4. Source Code Mengisi Missing Value
- Lampiran 5. Source Code One-hot Encoding
- Lampiran 6. Source Code Dataset Target
- Lampiran 7. Source Code Normalisasi Menggunakan Standard Scaler
- Lampiran 8. Source Code Reduksi Dimensi Menggunakan PCA
- Lampiran 9. Source Code Reduksi Dimensi Menggunakan LDA
- Lampiran 10. Source Code Split Data
- Lampiran 11. Source Code Klasifikasi Random Forest
- Lampiran 12. Source Code Confusion Matrix
- Lampiran 13. Source Code AUC