

**OPTIMALISASI METODE K-MEANS CLUSTERING DENGAN  
PRINCIPAL COMPONENT ANALYSIS (PCA) PADA PERFORMA  
PEMAIN SEPAK BOLA DI LIGA PRIMER INGGRIS**

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

Disusun Oleh

**AHMAD AFFAN AFIFUDDIN**

**H92218039**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA  
2024**

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama : AHMAD AFFAN AFIFUDDIN  
NIM : H92218039  
Prodi : Matematika  
Angkatan : 2018

Dengan ini menyatakan bahwa hasil skripsi ini merupakan hasil karya sendiri dan benar keasliannya. Apabila ternyata di kemudian hari penelitian ini merupakan hasil plagiat atau penjiplakan atas karya orang lain, maka saya bersedia bertanggung jawab sekaligus menerima sanksi.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Surabaya, 19 Juni 2024



Ahmad Affan Afifuddin

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : AHMAD AFFAN AFIFUDDIN  
NIM : H92218039  
Judul Skripsi : Optimalisasi Metode *K-Means Clustering* Dengan  
*Principal Component Analysis (PCA)* Pada Performa  
Pemain Sepak Bola Di Liga Primer Inggris

Telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Nurissaidah Ulinuha, M.Kom  
NIP. 199011022014032004

Pembimbing II



Aris Fanani, M.Kom  
NIP. 198701272014031002

Mengetahui,

Ketua Program Studi Matematika

UIN Sunan Ampel Surabaya



Yuniar Farida, M.T  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI

Skripsi oleh

Nama : AHMAD AFFAN AFIFUDDIN  
NIM : H92218039  
Judul Skripsi : Optimalisasi Metode *K-Means Clustering* Dengan  
*Principal Component Analysis (PCA)* Pada Performa  
Pemain Sepak Bola Di Liga Primer Inggris

Telah dipertahankan di depan Tim Penguji  
Pada tanggal 19 Juni 2024

Mengesahkan,  
Tim Penguji

Penguji I

Dr. Abdulloh Hamid, M.Pd  
NIP. 198508282014031003

Penguji II

Dian Yulianti, M.Si  
NIP. 198707142020122015

Penguji III

Nurissaidah Ulinuha, M.Kom  
NIP. 199011022014032004

Penguji IV

Aris Fanani, M.Kom  
NIP. 198701272014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya



Abdul Hamdani, M.Pd  
NIP. 196507312000031002

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : AHMAD AFFAN AFIFUDDIN  
NIM : H92218039  
Fakultas/Jurusan : Sains dan Teknologi / Matematika  
E-mail address : ahmadaffanmusic@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi  Tesis  Desertasi  Lain-lain (.....)  
yang berjudul :

OPTIMALISASI METODE K-MEANS CLUSTERING DENGAN PRINCIPAL COMPONENT

ANALYSIS (PCA) PADA PERFORMA PEMAIN SEPAK BOLA DI LIGA PRIMER INGGRIS

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 10 JULI 2024

Penulis



( AHMAD AFFAN AFIFUDDIN )  
*nama terang dan tanda tangan*



## ABSTRAK

### **Optimalisasi Metode K-Means Clustering Dengan Principal Component Analisis (PCA) Pada Performa Pemain Sepak Bola Di Liga Primer Inggris**

Klub sepak bola pasti memiliki keinginan yang luar biasa untuk memenangkan pertandingan dan juga trofi. Akibatnya, mereka akan mencari pemain yang terampil dan dapat mencetak gol, melakukan operan yang akurat, mengurani kesalahan, dan memahami konsep taktis. Penelitian ini bertujuan untuk mengklasifikasikan pemain berdasarkan performa di liga inggris menggunakan optimasi dari algoritma *K-Means* dengan metode *Principal Component Analysis (PCA)*. Metode *K-Means* memiliki beberapa keunggulan, antara lain di tingkat keakuratan dalam clustering cukup baik, mampu mengelompokkan data dengan kuantitas besar dalam waktu singkat, dan relatif mudah karena jumlah *cluster* dan pusat *cluster* dapat diinisialisasi secara acak. Data yang digunakan pada penelitian ini adalah statistik performa pemain sepak bola di Liga Primer Inggris untuk musim 2023-2024. Berdasarkan perhitungan, *cluster* yang ideal adalah  $k = 2$ , dengan 143 anggota dalam kategori *cluster* sangat baik dan 377 anggota dalam *cluster* baik. Cluster  $k = 2$  memiliki *Silhouet Coefisien (SC)* sebesar 0,632. Hal ini menandakan bahwa temuan *cluster* memiliki struktur baik.

**Kata kunci:** Sepak Bola, *Clustering*, *K-Means*, *Principal Component Analisis*

UIN SUNAN AMPEL  
S U R A B A Y A

## ABSTRACT

### **Optimization of K-Means Clustering Method with Principal Component Analysis (PCA) on Football Player Performance in English Premier League.**

Football clubs must have an overwhelming desire to win matches and also trophies. As a result, they will look for players who are skilled and can score goals, make accurate passes, reduce mistakes, and understand tactical concepts. This research aims to classify players by using optimization of the K-Means algorithm with the Principal Component Analysis (PCA) method. The K-Means method has several advantages, including the level of accuracy in clustering is quite good, able to group large quantities of data in a short time, and relatively easy because the number of clusters and cluster centers can be initialized randomly. The data used in this research is the performance statistics of soccer players in the English Premier League for the 2023-2024 season. Based on the calculation, the ideal cluster is  $k = 2$ , with 143 members in the excellent cluster category and 377 members in the good cluster. Cluster  $k = 2$  has a Silhouet Coefisien (SC) of 0.632. This indicates that the cluster findings have a good structure.

**Keywords:** Football, Clustering, K-Means, Principal Component Analysis

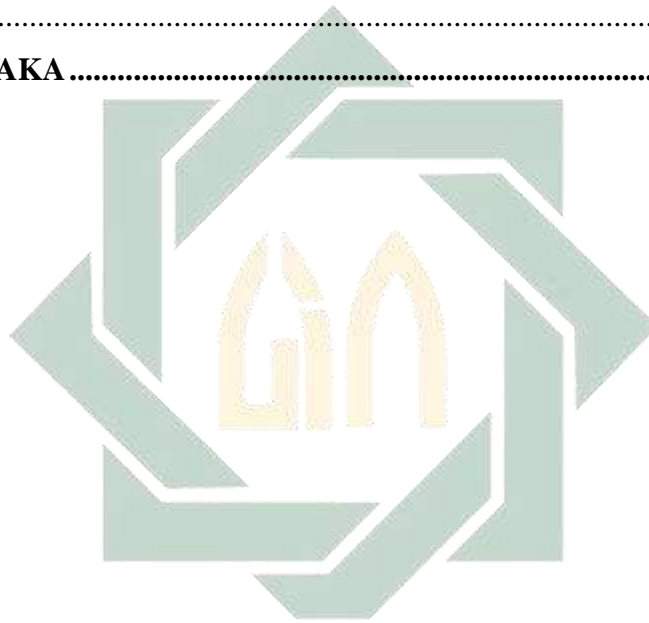
UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>PERNYATAAN KEASLIAN.....</b>	<b>i</b>
<b>LEMBAR PERSETUJUAN PEMBIMBING .....</b>	<b>ii</b>
<b>PENGESAHAN TIM PENGUJI .....</b>	<b>iii</b>
<b>LEMBAR PERSETUJUAN PUBLIKASI.....</b>	<b>iv</b>
<b>MOTTO .....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>viii</b>
<b>ABSTRACT .....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah .....	5
1.3. Tujuan Penelitian.....	6
1.4. Manfaat Penelitian.....	6
1.5. Batasan Masalah.....	6
1.6. Sistematika Penulisan.....	7
<b>KAJIAN PUSTAKA .....</b>	<b>8</b>
2.1. Sepak Bola .....	8
2.2. <i>Clustering</i> .....	9
2.3. <i>Uji Kaiser Meyer Olkin Measur of Sampling (KMO)</i> .....	10
2.4. <i>Principal Component Analysis (PCA)</i> .....	12
2.5. K-Means Clustering .....	14
2.6. Silhouette Coefficient .....	15
2.7. Integrasi Keilmuan.....	17
<b>METODE PENELITIAN .....</b>	<b>21</b>
3.1. Jenis Penelitian.....	21
3.2. Data Penelitian .....	21
3.3. Teknik Analisis Data.....	22
<b>HASIL DAN PEMBAHASAN .....</b>	<b>25</b>



4.1. Analisis Deskriptif Data .....	25
4.2. Pre-Processing Data.....	26
4.3. K-Means Clustering .....	29
4.4. Perbandingan Hasil.....	39
4.5. Integrasi Keilmuan .....	41
<b>BAB V.....</b>	<b>43</b>
<b>PENUTUP.....</b>	<b>43</b>
5.1. Simpulan .....	43
5.2. Saran .....	43
<b>DAFTAR PUSTAKA.....</b>	<b>44</b>



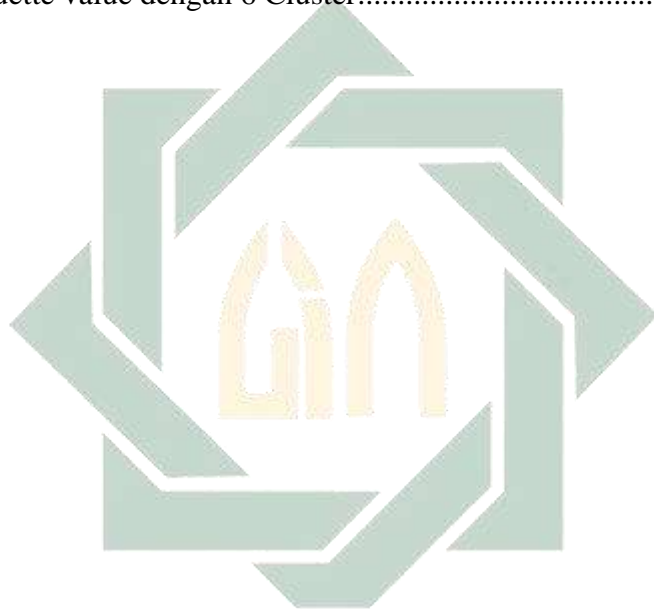
UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR TABEL

Tabel 2.1 Variabel Performa Pemain Sepak Bola.....	9
Tabel 2.2 Kriteria nilai uji KMO.....	12
Tabel 2.3 Silhouette Coefficient.....	17
Tabel 4.1 Data Performa Pemain Sepak Bola.....	25
Tabel 4.2 Deskripsi Data performa pemain sepak bola .....	25
Tabel 4.3 Hasil Uji KMO.....	26
Tabel 4.4 Matriks Kovarian .....	27
Tabel 4.5 Nilai Eigen .....	27
Tabel 4.6 Rotasi Faktor .....	28
Tabel 4.7 Data Performa Pemain Setelah Pre-Processing .....	29
Tabel 4.8 Centroid Awal .....	30
Tabel 4.9 Jarak Euclidean Iterasi ke-1 .....	30
Tabel 4.10 Centroid Baru Iterasi ke-2.....	31
Tabel 4.11 Jarak Euclidean Iterasi ke-2 .....	31
Tabel 4.12 Centroid Baru Iterasi ke-6.....	32
Tabel 4.13 Jarak Euclidian Iterasi Ke-6.....	32
Tabel 4.14 Centroid Baru Iterasi Ke-13.....	33
Tabel 4.15 Jarak Euclidian Iterasi Ke-13.....	33
Tabel 4.16 Centroid Baru Iterasi Ke-14.....	35
Tabel 4.17 Jarak Euclidian Iterasi Ke-14.....	35
Tabel 4.18 Centroid Baru Iterasi Ke-11.....	36
Tabel 4.19 Jarak Euclidian Iterasi Ke-11 .....	36
Tabel 4.20 Centroid Baru Iterasi Ke-30.....	38
Tabel 4.21 Jarak Euclidian Iterasi Ke-30.....	31
Tabel 4.22 Perbandingan Keakuratan Hasil 2 Sampai 6 Cluster .....	31
Tabel 2.23 Centroid optimal pada Cluster $K=2$ .....	40
Tabel 2.24 Pemain Anggota dari $K=2$ .....	40

## DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian .....	24
Gambar 4.1 silhouette value dengan 2 Cluster.....	33
Gambar 4.2 silhouette value dengan 3 Cluster.....	34
Gambar 4.3 silhouette value dengan 4 Cluster.....	36
Gambar 4.4 silhouette value dengan 6 Cluster.....	37
Gambar 4.5 silhouette value dengan 6 Cluster.....	39



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

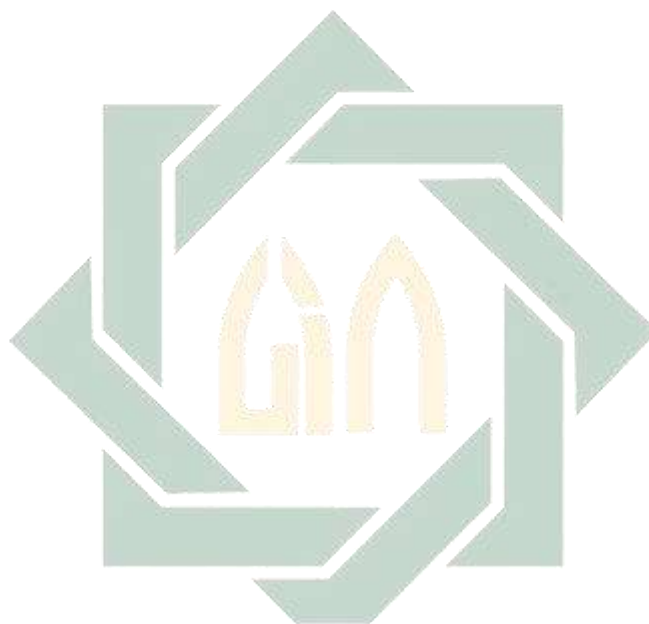
- Akbar, M. A., Fatimah, F., & Jaenudin, J. (2019, October). Penerapan Data Mining Untuk Pengelompokan Posisi Pemain Sepak Bola Menggunakan Algoritma K-Means Clustering. In *Seminar Nasional Teknologi Informasi* (Vol. 2, pp. 278-282).
- Annur, H. (2019). Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *Jurnal Informatika Upgris*, 5(1).
- Anwar, K., Goejantoro, R., and Prangga, S. (2022). Pengelompokan Kabupaten/Kota Di Pulau Kalimantan Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2020 Menggunakan Optimasi K-Means Cluster Dengan Principle Component Analysis (PCA) Grouping Of Regencies/Cities in Kalimantan Island Based On Human Development Index Indicators In 2020 Using K-Means Cluster Optimization Using Principle Component Analysis (PCA). 13:131–140.
- Atma, H., Putra, A. R., and Enri, Ultach, N. (2021). Clustering Clustering Data Eskspor Buah-Buahan Berdasarkan Negara Tujuan Menggunakan Algoritma KMeans. *Bina Insani Ict Journal*, 8(1):73
- D'Amado, L. J. (2020). Montevideo 1930: reassessing the selection of the first World Cup host. *Soccer and Society*, 00(00):848–860.
- Dwididanti, S., Anggoro, D. A., & Sutanto, M. H. (2022). Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit. *Emitor: Jurnal Teknik Elektro*, 22(2), 110-117.
- Farida, Y., Fahmi Khariri, A., Yuliati, D., Khaulasari, H., Sunan Ampel, U., Farida, Y., Khariri, A., Yuliati, D., and Khaulasari, H. (2022). Clustering Couples of Childbearing Age to Get Family Planning Counseling Using K-Means Method. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 22(1):189–200.
- Garcia-del Barrio, P. and Pujol, F. (2020). Recruiting talent in a global sports market: appraisals of soccer players' transfer fees. *Managerial Finance*, 47(6):789–811.

- Hartono, W., Hadi, S., Rosnawati, R., and Retnawati, H. (2022). Uji Kecocokan Model Parameter Logistik Soal Diagnosa Kemampuan Matematika Dasar. *JNPM (Jurnal Nasional Pendidikan Matematika)*, 6(1):125.
- Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *JIEET (Journal of Information Engineering and Educational Technology)*, 5(2), 49-54.
- <https://soccer.indozone.id/preview-review/982788268/3-transfer-mahal-liga-inggris-yang-flop-ada-pemain-mu-dan-chelsea>. Diakses pada 20 november 2023
- <https://www.idntimes.com/sport/soccer/audi-rahmantio-1/penyerang-mahal-yang-flop-di-manchester-united-c1c2>. Diakses pada 20 November 2023
- <https://www.indosport.com/sepakbola/20231113/6-transfer-aneh-manchester-united-yang-dipermasalahan-sir-jim-ratcliffe>. Diakses pada 20 November 2023.
- IFAB (2023). Laws of the Game 22/23. page 230.
- Jadecivicius, A. (2019). FA Cup, UK GDP, FTSE Growth and National House Prices. Available at SSRN 3334659.
- Jo, J.-M. (2019). Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance. *The Journal of the Korea Institute of Electronic Communication Sciences*, 14(3):547–552.
- Li, Y., & Mateos, G. (2022). Networks of international football: communities, evolution and globalization of the game. *Applied Network Science*, 7(1), 59.
- Martiano, M., Zarlis, M., and Wage, S. (2019). Modification K-Means Model With Local Deviation Method To Improve The Accuracy In Forming Clusters.
- Muningsih, E., Hasan, N., & Sulistyio, G. B. (2020). Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia. *Bianglala Informatika*, 8(1), 58-62.
- Nurjoko, Dwirohayati, D., and Sudibyio, N. H. (2020). Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering. *Teknika*, 14(2):127–135.



- Nurjoko, Dwirohayati, D., and Sudiby, N. H. (2020). Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering. *Teknika*, 14(2):127–135.
- Oktaviana, E. (2022). Clustering bencana alam di Indonesia menggunakan algoritma K-Means.
- Polindi, M. and Farida, I. (2019). Model Rekrutmen, Seleksi, Pelatihan Dan Pengembangan, Penilaian Kinerja Dan Kompensasi Karyawan Dala Perspektif Manajemen Syari'ah. *Jurnal Aghinya Stiesnu Bengkulu*, 2(1):90–105.
- Putra, I. P. W. A., & Devi, S. (2021). Pengaruh Team Status, Biaya Transfer dan Biaya Gaji Terhadap Market Value Pemain Sepak Bola Profesional (Studi Empiris Pada Himpunan Sepak Bola Yang Masih Berkompetisi dalam Liga Inggris Periode 2018-2019). *JIMAT (Jurnal Ilmiah Mahasiswa Akuntansi) Undiksha*, 12(1), 902-911.
- Putriana, P., Suarna, N., & Prihartono, W. (2023). ANALISIS CLUSTERING PRESTASI ATLET PADA BERBAGAI CABANG OLAHRAGA MENGGUNAKAN ALGORITMA K-MEANS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3435-3442.
- Ritonga, A. S., & Muhandhis, I. (2021). Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (Pca). *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, 7(2), 124-133.
- Simanjuntak, K. P., & Khaira, U. (2021). Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering: Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 7-16.
- Suryadi, U. T., & Supriatna, Y. (2019). Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi dan Komunikasi*, 14(1), 15-27.
- Vafaei, N., Ribeiro, R. A., Camarinha-Matos, L. M., and Valera, L. R. (2020). Normalization techniques for collaborative networks. *Kybernetes*, 49(4):1285– 1304.

- Velema, T. A. (2021). Globalization and player recruitment: How teams from European top leagues broker migration flows of footballers in the global transfer network. *International Review for the Sociology of Sport*, 56(4):493–513.
- Wiyantoro, A. S. (2019). Seleksi Fitur dan Preferensi Penyerang Terbaik Liga Inggris Berbasis Fisher's Discriminant Ratio, K-Means Clustering dan Topsis. *JURNAL ILMIAH INFORMATIKA*, 7(02), 76-81.



UIN SUNAN AMPEL  
S U R A B A Y A