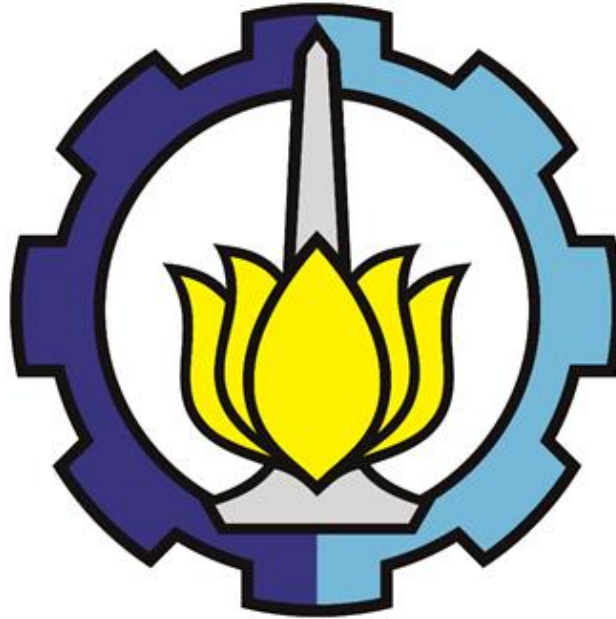


PROGRESS FINAL PROJECT

KECERDASAN KOMPUTASIONAL - E



Disusun Oleh:

Ersad Ahmad Ishlahudin	05111740000016
Bayu Laksana	05111740000020
Andhika Yoga Perdana	05111740000101

Dosen:

Dini Adni Navastara, S.Kom, M.Sc

DEPARTEMEN INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

2019

Deskripsi Dataset

Dataset Classification

Dataset yang kami gunakan untuk classification adalah dataset yang diambil dari situs UCI, dengan judul datasetnya adalah "MAGIC gamma telescope data 2004"

Konteks

Data "MAGIC gamma telescope data 2004" digunakan untuk mensimulasikan registrasi energi tinggi partikel gamma dalam teleskop Cherenkov atmosfer berbasis atmospheric Cherenkov gamma telescope menggunakan teknik pencitraan.

Dataset tersebut terdiri atas **10 atribut** dan **19020 data**.

Informasi Atribut

No	Atribut	Tipe	Deskripsi
1	fLength	Numeric	sumbu mayor elips [mm]
2	fWidth	Numeric	sumbu minor elips [mm]
3	fSize	Numeric	10-log jumlah konten dari semua piksel [in #phot]
4	fConc	Numeric	rasio jumlah dua piksel tertinggi yang yang lebih fSize [rasio]
5	fConc1	Numeric	rasio piksel tertinggi yang lebih dari fSize [rasio]
6	fAsym	Numeric	jarak dari piksel tertinggi ke pusat, diproyeksikan ke sumbu mayor [mm]
7	fM3Long	Numeric	Root ke-3 dari momen ketiga sepanjang sumbu mayor[mm]
8	fM3Trans	Numeric	Root ke-3 dari momen ketiga sepanjang sumbu minor[mm]
9	fAlpha	Numeric	sudut sumbu utama dengan vektor ke asal [deg]
10	fDist	Numeric	jarak dari asal ke pusat elips [mm]

Informasi Kelas

1	class	Kategorikal	Terdiri atas gamma (signal) dan hadron (background)
---	-------	-------------	---

Tautan: <https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope>

Dataset Clustering

Dataset yang kami gunakan untuk clustering adalah dataset yang diambil dari situs Kaggle, dengan judul datasetnya adalah "Predicting a Pulsar Star". Dataset dapat diakses melalui tautan yang disediakan di bawah.

Tautan : https://www.kaggle.com/pavanraj159/predicting-a-pulsar-star#pulsar_stars.csv

Deskripsi Dataset

Bintang pulsar adalah jenis bintang Neutron yang langka yang menghasilkan emisi radio yang dapat dideteksi di Bumi. Mereka sangat menarik secara ilmiah sebagai wahana ruang-waktu, medium antar-bintang, dan keadaan materi.

Setiap pulsar menghasilkan pola emisi yang sedikit berbeda, yang mana juga berbeda pada setiap putaran. Dengan demikian deteksi sinyal potensial yang dikenal sebagai 'kandidat', dirata-ratakan pada banyak rotasi pulsar, sebagaimana ditentukan oleh panjang pengamatan.

Dengan tidak adanya info tambahan, masing-masing kandidat berpotensi menggambarkan pulsar yang sebenarnya. Namun dalam praktiknya hampir semua deteksi disebabkan oleh interferensi frekuensi radio (RFI) dan noise, yang membuat sinyal yang valid sulit ditemukan.

Machine learning sekarang digunakan untuk memberikan label (mengklasifikasikan) kandidat bintang pulsar secara otomatis untuk mempercepat proses analisis. Sistem klasifikasi sedang banyak diadopsi, yang memperlakukan dataset sebagai permasalahan binary classification.

Dataset yang berisi 16.259 contoh palsu yang disebabkan oleh RFI / noise, dan 1.639 contoh pulsar nyata. Label kelas yang digunakan adalah 0 (negatif) dan 1 (positif).

Informasi Atribut

Setiap kandidat dipaparkan oleh 8 atribut kontinu, dan satu kolom label (kelas). Empat atribut yang pertama adalah statistik sederhana yang diperoleh dari profil pulsa terintegrasi. Empat variabel sisanya diperoleh dengan cara yang sama dari kurva DM-SNR.

No	Atribut	Nilai
1	Mean of the integrated profile	Numerik
2	Standard deviation of the integrated profile	Numerik
3	Excess kurtosis of the integrated profile	Numerik
4	Skewness of the integrated profile	Numerik
5	Mean of the DM-SNR curve	Numerik
6	Standard deviation of the DM-SNR curve	Numerik

7	Excess kurtosis of the DM-SNR curve	Numerik
8	Skewness of the DM-SNR curve	Numerik

Metode dan Skenario

Classification + Optimasi

Algoritma yang digunakan:

- K-Nearest Neighbor
- Untuk optimasi menggunakan Genetic Algorithm

Metode Evaluasi: Confusion Matrix

Skenario :

Kelompok kami akan membandingkan pemrosesan data "MAGIC gamma telescope data 2004" dengan menggunakan algoritma K-Nearest Neighbor (tanpa menggunakan metode optimasi) dan algoritma K-Nearest Neighbor yang telah diimplementasikan dengan metode optimasi yaitu Genetic Algorithm. Parameter atau variabel dalam algoritma K-NN yang dapat diubah yaitu nilai K (jumlah tetangga) dan persentase antara data test dan data training.

Clustering + Optimasi

Algoritma clustering yang digunakan :

- K-Means
- Optimasi menggunakan Genetic Algorithm

Metode evaluasi performance:

- Silhouette Coefficient
- Dunn index

Skenario :

Skenario yang hendak kami terapkan pada clustering dan optimasi pada dataset "Predicting a Pulsar Star" adalah dengan melakukan komparasi (perbandingan) algoritma K-Means tanpa menggunakan optimasi dengan algoritma K-Means yang telah diterapkan optimasi menggunakan optimasi Genetic Algorithm. Optimasi pada algoritma K-Means dapat diterapkan pada parameter berupa K (atau jumlah cluster) atau inisialisasi posisi centroid.

Referensi

Paper

- https://www.researchgate.net/profile/Bashar_Al-Shboul/publication/258332160_Initializing_K-Means_using_Genetic_Algorithms/links/590cb050aca2722d185c00c2/Initializing-K-Means-using-Genetic-Algorithms.pdf?origin=publication_detail
- <https://www.sciencedirect.com/science/article/pii/S2212017313004945>
- https://www.ripublication.com/ijaer17/ijaerv12n24_35.pdf
- https://res.mdpi.com/d_attachment/algorithms/algorithms-10-00123/article_deploy/algorithms-10-00123.pdf