

## INTISARI

### ANALISIS KLASIFIKASI DENGAN REBMIX-EXPECTATION-MAXIMIZATION MELALUI GAUSSIAN MIXTURE MODEL

Oleh

FIRDA ROKHANA ARFIATUNNISWA

17/409517/PA/17824

Analisis klasifikasi melalui *Gaussian Mixture Model* merupakan analisis klasifikasi yang mengasumsikan setiap kelas merupakan *Gaussian mixture* dari setiap subkelas. Algoritma *Expectation-Maximization* (EM) digunakan untuk memperoleh estimasi parameter *Gaussian Mixture Model* berupa bobot campuran, vektor mean, dan matriks kovariansi setiap subkelas. Algoritma *Rough-Enhanced-Bayes Mixture Estimation* (REBMIX) digunakan untuk menentukan nilai inisialisasi bagi algoritma EM untuk meningkatkan hasil estimasi dan mempercepat proses estimasi. Penentuan observasi masuk ke kelas satu atau nol didasarkan pada *Log Likelihood Ratio Test* (LLRT) dengan nilai ambang batas tes sesuai *False Accept Rate* (FAR) yang ditentukan.

Analisis klasifikasi melalui *Gaussian Mixture Model* dengan estimasi parameter menggunakan algoritma REBMIX-EM yang digunakan pada data Diabetes menghasilkan nilai *Area Under Curve* (AUC) kurva *Receiver Operating Characteristic* (ROC) sebesar 0,953 yang menunjukkan model klasifikasi sangat baik dalam memisahkan kelas satu dan kelas nol. Hasil analisis klasifikasi menunjukkan semakin besar nilai *False Accept Rate* (FAR) maka semakin besar nilai *Apparent Error Rate* (APER) dan sensitivitas, serta semakin kecil nilai akurasi dan spesifisitas yang dihasilkan. Analisis klasifikasi dengan algoritma REBMIX-EM melalui *Gaussian Mixture Model* menghasilkan performa klasifikasi yang lebih baik dibanding Regresi Logistik dan Naive Bayes.

Kata kunci: Analisis klasifikasi, Algoritma *Expectation-Maximization*, Algoritma *Rough-Enhanced-Bayes Mixture Estimation*, *Gaussian Mixture Models*, *Log Likelihood Ratio Test*

## ABSTRACT

### **REBMIX- EXPECTATION-MAXIMIZATION CLASSIFICATION BY GAUSSIAN MIXTURE MODEL**

By

FIRDA ROKHANA ARFIATUNNISWA

17/409517/PA/17824

Classification analysis through the Gaussian Mixture Model is a classification analysis that assumes each class is a Gaussian mixture of each subclass. The Expectation-Maximization (EM) algorithm is used to estimate parameters of the Gaussian Mixture Model, including mixture weights, mean vectors, and covariance matrices for each subclass. The Rough-Enhanced-Bayes Mixture Estimation (REBMIX) algorithm is used to determine the initial values for the EM algorithm to improve estimation results and speed up the estimation process. The determination of whether an observation belongs to class one or zero is based on the Log Likelihood Ratio Test (LLRT) with the test threshold value corresponding to the specified False Accept Rate (FAR).

Classification analysis through the Gaussian Mixture Model with parameter estimation using the REBMIX-EM algorithm applied to Diabetes data produced an Area Under Curve (AUC) value of 0.953 on the Receiver Operating Characteristic (ROC) curve, indicating that the classification model is very good at separating class one and class zero. The classification analysis results show that the larger the False Accept Rate (FAR), the larger the Apparent Error Rate (APER) and sensitivity, and the smaller the accuracy and specificity. Classification analysis with the REBMIX-EM algorithm through the Gaussian Mixture Model produces better classification performance compared to Logistic Regression and Naive Bayes.

**Keywords:** Classification Analysis, Expectation-Maximization Algorithm, Rough-Enhanced-Bayes Mixture Estimation Algorithm, Gaussian Mixture Models, Log Likelihood Ratio Test