

INTISARI

ANALISIS DISKRIMINAN BAYESIAN DENGAN METODE *MINIMUM EXPECTED COST OF MISCLASSIFICATION* (ECM) DAN *DECISION TREE* PADA DATA PENDERITA DIABETES

Oleh:

Awang Putra Sembada R

16/394189/PA/17280

Terdapat banyak metode pengklasifikasian, salah satunya adalah Analisis Diskriminan Bayesian dengan Minimum *Expected Cost of Misclassification* (ECM). Metode ini merupakan analisis diskriminan yang dalam analisisnya memperhatikan biaya kesalahan klasifikasi. Namun, sebelum dilakukan analisis perlu memenuhi beberapa asumsi. Asumsi utamanya yaitu kesamaan matriks varian kovarian tiap kelas dan variabel independen berdistribusi normal multivariat tiap kelas. Dalam mendapatkan rumus diskriminan metode Minimum *Expected Cost of Misclassification* (ECM) dengan bantuan teorema Bayes yang tidak lepas dari probabilitas posterior. Untuk melihat performa dari metode klasifikasi ini dilakukan dengan pembandingan metode lainnya yaitu metode *decision tree*. Data yang digunakan merupakan data penderita diabetes. Berdasarkan beberapa variabel independen suatu individu akan masuk ke kelas negatif atau positif diabetes. Dilakukan transformasi Tukey sebelum analisis diskriminan agar asumsi normalitas multivariat dan kesamaan matriks varian kovarian dapat dipenuhi. Untuk tingkat kesalahan diukur dengan melihat *Apparent Error Rate* (APER), akurasi, presisi, sensitivitas, dan spesifisitas. Hasil dari analisis menunjukkan diskriminan Bayesian metode minimum *Expected Cost of Misclassification* (ECM) setelah transformasi Tukey lebih baik dibandingkan dengan metode *decision tree* berdasarkan APER, akurasi, presisi, dan sensitivitas.

Kata Kunci: Analisis Diskriminan Bayesian, Teorema Bayes, Analisis Diskriminan Linear, *Decision Tree*, Transformasi Tukey, Analisis Diskriminan Metode Minimum *Expected Cost of Misclassification* (ECM)

ABSTRACT

BAYESIAN DISCRIMINANT ANALYSIS WITH MINIMUM EXPECTED COST OF MISCLASSIFICATION (ECM) AND DECISION TREE METHOD ON DIABETES PATIENT DATA

By:

Awang Putra Sembada R

16/394189/PA/17280

There are many classification methods, one of which is Bayesian Discriminant Analysis with Minimum Expected Cost of Misclassification (ECM). This method is a discriminant analysis which in its analysis considers the cost of misclassification. However, before the analysis is carried out, it is necessary to fulfill several assumptions. The main assumption is the similarity of the variance covariance matrix for each class and the independent variables has a multivariate normal distribution for each class. To get the discriminant formula the Minimum Expected Cost of Misclassification (ECM) method with the help of the Bayes theorem which cannot be separated from posterior probability. To see the performance of this classification method, a comparison method is used, namely the decision tree method. The data used is data on diabetics. The goal is that based on several independent variables an individual will enter the negative or positive class for diabetes. The Tukey transformation is carried out before discriminant analysis so that the assumptions of multivariate normality and variance covariance matrix equality can be fill. The error rate is measured by looking at the Apparent Error Rate (APER), accuracy, precision, sensitivity, and specificity. The results of the analysis show that the Bayesian discriminant method of minimum Expected Cost of Misclassification (ECM) after the Tukey transformation is better than the decision tree method based on APER, accuracy, precision, and sensitivity.

Keywords: Bayesian Discriminant Analysis, Bayes Theorem, Linear Discriminant Analysis, Decision Tree, Tukey Transform, Discriminant Analysis Minimum Expected Cost of Misclassification (ECM) Method