

INTISARI

PERBANDINGAN METODE ANALISIS DISKRIMINAN LINIER DAN REGRESI LOGISTIK MULTINOMIAL UNTUK KLASIFIKASI PADA DATA MULTIKELAS

Oleh

Muhammad Tahta Rizqi Yuandri

20/462318/PA/20290

Proses klasifikasi merupakan salah satu proses analisis data yang dapat digunakan untuk mengklasifikasikan sebuah objek berdasarkan kriteria atau ciri-ciri yang telah dimiliki dari objek tersebut sehingga dengan kriteria yang telah dimiliki dapat dilakukan proses klasifikasi ke dalam kelas yang sesuai. Pada skripsi ini dibahas mengenai perbandingan metode klasifikasi, yaitu dengan menggunakan metode regresi logistik multinomial dan analisis diskriminan linier. Dari kedua metode tersebut akan dicari akurasi dan nilai *f1-score* terbaik sehingga dapat direkomendasikan untuk penelitian selanjutnya. Data yang digunakan adalah data klasifikasi jenis penyakit obesitas.

Sebelum melakukan analisis regresi logistik multinomial dan analisis diskriminan linier, dilakukan uji asumsi yang diperlukan dalam penelitian ini, yaitu uji normalitas multivariat, uji homogenitas matriks varian kovarian, dan uji nilai vektor rata-rata. Selanjutnya akan dilakukan analisis dengan menggunakan kedua metode yang telah disebutkan sebelumnya. Dilakukan *partitioning data* untuk membagi menjadi *data training* dan *data testing*. Dengan menggunakan *data training* akan dilakukan pemodelan regresi logistik multinomial dan analisis diskriminan linier. Setelah mendapatkan masing-masing model, akan dilakukan prediksi pada *data testing* dan dilakukan penghitungan akurasi untuk setiap metode yang digunakan. Pada tabel akurasi terdapat berapa jumlah data yang salah untuk

diklasifikasikan dan terdapat juga jumlah data yang tepat untuk diklasifikasikan sesuai dengan kelasnya. Didapatkan akurasi metode analisis regresi logistik multinomial sebesar 93.04% dan akurasi pada metode analisis diskriminan linier sebesar 88.61% serta didapatkan nilai *f1-score* untuk setiap kelas lebih baik pada metode regresi logistik multinomial daripada analisis diskriminan linier. Selain itu, setelah dilakukan reduksi variabel prediktor, tingkat akurasi regresi logistik multinomial masih baik dibandingkan dengan analisis diskriminan. Oleh karena itu, pada penelitian kali ini dengan menggunakan data klasifikasi jenis penyakit obesitas didapatkan metode yang lebih cocok untuk digunakan adalah metode regresi logistik multinomial.

Kata Kunci: Regresi logistik multinomial, analisis diskriminan linier, analisis klasifikasi, tingkat akurasi, *f1-score*.

ABSTRACT

COMPARISON OF LINEAR DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION METHODS FOR CLASSIFICATION OF MULTICLASS DATA

By

Muhammad Tahta Rizqi Yuandri

20/462318/PA/20290

The classification process is one of the data analysis processes that can be used to classify an object based on criteria or characteristics that have been owned from the object so that with the criteria that have been owned the classification process can be carried out into the appropriate class. This thesis discusses the comparison of classification methods, namely by using the multinomial logistic regression method and linear discriminant analysis. From both methods, the best accuracy and f1-score value will be sought so that it can be recommended for further research. The data used is obesity disease type classification data.

Before performing multinomial logistic regression analysis and linear discriminant analysis, the assumption tests required in this study are carried out, namely the multivariate normality test, the covariance variance matrix homogeneity test, and the mean vector value test. Next, the analysis will be carried out using the two methods mentioned earlier. Data partitioning is done to divide into training data and testing data. Using the training data, multinomial logistic regression modeling and linear discriminant analysis will be performed. After obtaining each model, predictions will be made on the testing data and accuracy calculations for each method used. In the accuracy table there is how much data is wrong to be classified and there is also the right amount of data to be classified according to its class. The accuracy of the multinomial logistic regression analysis

method is 93.04% and the accuracy of the linear discriminant analysis method is 88.61% and the f1-score value for each class is better in the multinomial logistic regression method than linear discriminant analysis. In addition, after reducing the predictor variables, the accuracy of multinomial logistic regression is still good compared to discriminant analysis. Therefore, in this study using obesity disease type classification data, it was found that the more suitable method to use was the multinomial logistic regression method.

Keywords: *Multinomial regression logistic, linear discriminant analysis, classification analysis, level of accuracy, f1-score.*