

ABSTRAK

Risma Noviana, Luh Putu (2022), Analisis Kinerja *Naïve Bayes Classifier* Berdasarkan Seleksi Fitur Berbasis *Information Gain* Dengan Seleksi Fitur Berbasis Multikolinearitas (Pada Studi Kasus Stunting Balita Di Puskesmas Kintamani). Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I: Dr. I Gede Aris Gunadi, S.Si. M.Kom dan Pembimbing II: Dr. I Made Gede Sunarya, S.Kom., M.Cs

Kata-kata kunci: Stunting, *Naïve Bayes Classifier*, *Information Gain*, Multikolinearitas, *Confusion Matrix*, Uji Koefisien Determinasi (*R Square*)

Penelitian ini bertujuan untuk mengetahui kinerja *Naïve Bayes Classifier* dengan membandingkan beberapa seleksi fitur. Pertama digunakan metode seleksi fitur berbasis *Information Gain*, kedua digunakan metode seleksi fitur berbasis Multikolinearitas, serta ketiga metode seleksi fitur yang merupakan kombinasi *Information Gain* dan Multikolinearitas. Dataset yang digunakan merupakan data kasus stunting balita di Puskesmas Kintamani I dan VI sebanyak 337 data. Dalam pengujian digunakan 4 skenario pengujian: (1) Skenario pengujian pertama, pada dataset langsung diimplementasikan algoritma *Naïve Bayes Classifier* dengan data training sebanyak 70% dan testing 30%, selanjutnya ditentukan nilai akurasi dan persisinya. (2) Skenario pengujian kedua, pada dataset langsung diimplementasikan algoritma *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain* dengan data training sebanyak 70% dan testing 30%, selanjutnya ditentukan nilai akurasi dan persisinya. (3) Skenario pengujian ketiga, pada dataset langsung diimplementasikan algoritma *Naïve Bayes Classifier* seleksi fitur Multikolinearitas dengan data training sebanyak 70% dan testing 30%, selanjutnya ditentukan nilai akurasi dan persisinya. (4) Skenario pengujian keempat, pada dataset langsung diimplementasikan algoritma *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain* dan fitur Multikolinearitas dengan data training sebanyak 70% dan testing 30%, selanjutnya ditentukan nilai akurasi dan persisinya. Sebagai tambahan juga dilakukan uji koefisien determinasi (*R Square*) untuk mengetahui pengaruh yang diberikan variabel bebas atau variabel Independent terhadap variabel terikat atau variabel dependent untuk memprediksi dan melihat seberapa besar kontribusi pengaruh yang diberikan terhadap variabel bebas secara simultan terhadap variabel terikat. Tahap pengujian yang digunakan dari 4 skenario yaitu *confusion matrix*. Sehingga diperoleh hasil akurasi dari masing-masing metode digunakan yaitu: *Naïve Bayes Classifier* sebesar 90,10%, *Naïve Bayes Classifier* dengan fitur *Information Gain* sebesar 95,05%, *Naïve Bayes Classifier* dengan fitur Multikolinearitas sebesar 93,07% dan *Naïve Bayes Classifier* dengan fitur *Information Gain* dan Multikolinearitas sebesar 96,04%. Maka dari 4 skenario yang digunakan metode *Naïve Bayes*

Classifier dengan fitur *Information Gain* dan Multikolinearitas mendapatkan hasil akurasi paling tinggi. Sehingga dilakukan uji koefisien determinasi (R Square) terhadap fitur *Information Gain* dan Multikolinearitas dengan menggunakan tools SPSS sebesar 0,577, hal ini berarti 57,7% semua fitur berpengaruh terhadap klasifikasi stunting. Dengan fitur variabel independen seperti umur, BPJS, imunisasi, merokok, riwayat ibu hamil, sementara sisanya sebesar 42,3% dipengaruhi faktor lainnya yang tidak dibahas dalam penelitian ini.



ABSTRACT

Performance Analysis of Naïve Bayes Classifier Based on Information Gain-Based Feature Selection With Multicollinearity Based Feature Selection (In the Case Study of Toddler Stunting at the Kintamani Health Center). Thesis, Computer Science, Graduate Program, Ganesha University of Education.

This thesis has been supervised and approved by Supervisor I: Dr. I Gede Aris Gunadi, S.Si. M.Kom and Supervisor II: Dr. I Made Gede Sunarya, S.Kom., M.Cs.

Keywords: Stunting, Naïve Bayes Classifier, Information Gain, Multikolinearitas, Confussion Matri, Coefficient of Determination Test (R Square).

The purpose of this research is to determine the performance of the Naïve Bayes Classifier by comparing several feature selection methods. The first is the Information Gain-based feature selection method, the second is the Multicollinearity-based feature selection method, and the third is the feature selection method which is a combination of Information Gain and Multicollinearity. The dataset used is data on toddler stunting cases at the Kintamani I and VI Health Centers totaling 337 data. In testing, 4 test scenarios were used: (1) In the first test scenario, the Naïve Bayes Classifier algorithm was directly implemented on the dataset with 70% training data and 30% testing, then the accuracy and precision values were determined. (2) The second test scenario, the Naïve Bayes Classifier algorithm is directly implemented on the dataset with Information Gain feature selection. With 70% training data and 30% testing, then the accuracy and precision values are determined. (3) The third test scenario, the Naïve Bayes Classifier algorithm directly implements the Multicollinearity feature selection on the dataset. With 70% training data and 30% testing, then the accuracy and precision values are determined. (4) The fourth test scenario, the Naïve Bayes Classifier algorithm is directly implemented on the dataset with Information Gain feature selection and Multicollinearity features. With 70% training data and 30% testing, then the accuracy and precision values are determined. In addition, a test of the coefficient of determination (R Square) was also carried out to determine the effect of the independent variable or independent variable on the dependent variable or variable to predict and see how much influence it contributed to the independent variable simultaneously on the dependent variable. The testing phase used from 4 scenarios is the confusion matrix. In order to obtain the results of the accuracy of each method used, namely: Naïve Bayes Classifier of 90.10%, Naïve Bayes Classifier with Information Gain feature of 95.05%, Naïve Bayes Classifier with Multicollinearity feature of 93.07% and Naïve Bayes Classifier with Information Gain and Multicollinearity features of 96.04%. So from the 4 scenarios used the Naïve Bayes Classifier method with the Information Gain and Multicollinearity features got the highest accuracy results. So a test of the coefficient of determination (R Square) was carried out on the Information Gain and Multicollinearity features using SPSS

tools of 0.577, this means that 57.7% of all features affect the stunting classification. With independent variable features such as age, BPJS, immunization, smoking, history of pregnant women, while the remaining 42.3% is influenced by other factors not discussed in this study.

