

MUHAMMAD ZEIN FATAHILLAH, 20.240.0130

PENERAPAN DATA MINING MENGGUNAKAN METODE NAIVE BAYES  
UNTUK PENENTUAN PENYAKIT DIABETES, di bawah bimbingan

Arif Soma Dermawan, M.Kom, Nurul Amalia, M.Kom

75 + xiii halaman/ 12 gambar/ 12 tabel/3 lampiran/25 pustaka

**ABSTRAK**

Diabetes merupakan penyakit gangguan metabolik menahun akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Untuk memudahkan pendektasian dini, maka penelitian ini akan membuat sebuah sistem dengan menerapkan algoritma naive bayes yang dapat membantu dalam memprediksi penyakit diabetes pada individu dengan akurat dan efektif. Pada penelitian ini menggunakan dataset dari Institusi Nasional Diabetes dan Penyakit Pencenaan dan Ginjal yang sudah di publik dengan nama alamat websitenya yaitu <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-> dataset dengan jumlah data sebanyak 768 orang dengan 9 atribut. Atributnya terdiri dari *pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, age* dan *outcome*. Metode pengembangan penelitian ini menggunakan CRISP-DM yang terdiri dari *business understanding, data understanding, data preparation, modelling, evaluation, dan deployment*. Metode pengujian sistem menggunakan blackbox, UAT dan confusion matrix. Hasil penelitiannya  $accuracy=79.2\%$ ,  $recall=67.4\%$ ,  $precision=61.7\%$ , dan  $f-1\ score=64.4\%$  dan sistem dengan menerapkan algoritma naive bayes berjalan dengan baik. Pada penelitian ini algoritma naive bayes mudah diimplementasikan, memberikan solusi klasifikasi yang cukup cepat untuk dataset yang cukup besar, dan bekerja secara efisien untuk dataset yang memiliki banyak atribut atau dimensi tinggi. Namun algoritma naive bayes memberikan estimasi probabilitas yang tidak sesuai dengan data, jika terdapat outlier dalam dataset dan tidak memberikan hasil yang optimal, jika dataset yang memiliki atribut yang bersifat kontinu seperti BMI, Diabetes Pedigree Function.

**Kata Kunci** : Diabetes, Data Mining, Algoritma Naive Bayes, Prediksi Probabilitas

MUHAMMAD ZEIN FATAHILLAH, 20.240.0130

PENERAPAN DATA MINING MENGGUNAKAN METODE NAIVE BAYES  
UNTUK PENENTUAN PENYAKIT DIABETES, di bawah bimbingan

Arif Soma Dermawan, M.Kom, Nurul Amalia, M.Kom

75 + xiii halaman/ 12 gambar/ 12 tabel/ 3 lampiran/ 25 pustaka

**ABSTRAK**

*Diabetes is a disorder metabolic disease as a result of the pancreas does not produce enough insulin or the body cannot use the insulin produced effectively. To facilitate early detection, this research will create a system by applying the naive bayes algorithm that can help predict diabetes in individuals accurately and effectively. In this study using datasets from the National Institute of Diabetes and Kidney and Kidney Diseases that have been made public with the name of the website <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>, namely <https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>. The attributes consist of pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, age and outcome. The development method of this research uses CRISP-DM which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The testing method uses blackbox, UAT and confusion matrix. Results of the research are accuracy = 79.2%, recall = 67.4%, precision = 61.7%, and f-1 score = 64.4% and the system by applying the naive bayes algorithm runs well. In this study, the naive bayes algorithm is easy to implement, provides a fairly fast classification solution for a large enough dataset, and works efficiently for datasets that have many attributes or high dimensions. However algorithm naive bayes gives probability estimates that do not fit the data, if there are outliers in the dataset and does not provide optimal results, if the dataset has continuous attributes such as BMI, Diabetes Pedigree Function.*

**Keyword** *Keyword : Diabetes, Data Mining, Algorithm Naive Bayes, Prediction Probability*