

**ANALISIS PERBANDINGAN METODE *K NEAREST
NEIGHBOR* DAN *GAUSSIAN NAIVE BAYES* PADA
KLASIFIKASI JURUSAN SISWA**

(Studi Kasus pada Siswa SMA Muhammadiyah 3 Jember)

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
Herdian Cahyaningrum, Deni Arifianto, Ginanjar Abdurrahman

ABSTRAK

Penjurusan siswa merupakan tugas wajib bagi sekolah menengah atas khususnya SMA Muhammadiyah 3 Jember dalam menentukan minat tiap siswanya. Penelitian ini akan mengupas tentang klasifikasi penjurusan siswa menggunakan metode *Gaussian Naive Bayes* dan *K Nearest Neighbor*. Data dalam penelitian ini menggunakan data siswa SMA Muhammadiyah 3 Jember dengan jumlah total 343 siswa. Fitur yang digunakan dalam penjurusan siswa adalah rekapitulasi nilai ujian nasional bahasa Indonesia, nilai ujian nasional bahasa inggris, nilai ujian nasional IPA, nilai ujian nasional matematika, ulangan harian IPA, Matematika, IPS, Bahasa Indonesia, Bahasa inggris, nilai tes verbal linguistik, logis matematis, spasial, kinestik, musical, interpersonal, intrapersonal dan natural. Terdapat 23 data yang tidak memiliki *output* dalam data ini sehingga total data yang dihitung berjumlah 320 data. Dalam *preprocessing* data digunakan metode *Median Substitution* untuk mengatasi permasalahan *Missing Value* dan metode *Min-Max Normalization* untuk menormalisasi data. Untuk mengatasi ketidakseimbangan data pada penelitian ini menggunakan metode *SMOTE (Synthetic Minority Oversampling Technique)*. Dari data sintetis hasil *SMOTE* diperoleh data total 381 data. Skenario uji dalam penelitian ini menggunakan metode *K fold Cross Validation* dengan nilai *k fold* = 2, 4, 5, 8 dan 10. Dalam pengukuran jarak, vektor yang digunakan dalam implementasi *K Nearest Neighbor* menggunakan *Euclidean Distance*. Hasil akurasi tertinggi metode *Gaussian Naive Bayes* adalah 83,33% sedangkan akurasi tertinggi yang diperoleh metode *K Nearest Neighbor* adalah 83,61% dengan nilai *k neighbor* = 7. Dalam penelitian ini metode *K Nearest Neighbor* lebih baik dalam melakukan klasifikasi penjurusan pada data siswa SMA Muhammadiyah 3 Jember.

Keyword : Klasifikasi, *Gaussian Naive Bayes*, *K-NN*, Penjurusan siswa, *SMOTE*, *Cross Fold Validation*.

**A Comparative Analysis of K Nearest Neighbor and Gaussian Naive Bayes
methods in students major classification
(A Case Study In SMA Muhammadiyah 3 Jember)**

*Department of Informatics Engineering Faculty of Engineering, University of
Muhammadiyah Jember*

Herdian Cahyaningrum, Deni Arifianto, Ginanjar Abdurrahman

ABSTRACT

Majoring students is a mandatory task for senior high schools, especially SMA Muhammadiyah 3 Jember in determining the interests of each student. This study will explore the classification of students' majors using the Gaussian Naive Bayes and K Nearest Neighbor methods. The data in this study used data from high school students of Muhammadiyah 3 Jember with a total of 343 students. The features used in the student majors are recapitulation of Indonesian national exam scores, English national exam scores, science national exam scores, math national exam scores, science daily tests, Mathematics, Social Studies, Indonesian, English, verbal linguistics test scores, logical mathematical, spatial, kinesthetic, musical, interpersonal, intrapersonal and natural. There are 23 data that do not have output in this data so that the total data calculated is 320 data. In preprocessing the data, the Median Substitution method is used to overcome the problem of Missing Value and the method Min-Max Normalization to normalize data. To overcome the imbalance of data in this study using the SMOTE (Synthetic Minority Oversampling Technique) method. From the synthetic data of SMOTE results obtained a total of 381 data. The test scenario in this study uses the K fold Cross Validation method with k fold values = 2, 4, 5, 8 and 10. In measuring distance, the vector used in the implementation of K Nearest Neighbor uses Euclidean Distance . The highest accuracy result of the Gaussian Naive Bayes method is 83.33% while the highest accuracy obtained by the K Nearest Neighbor method is 83.61% with a value of k neighbor = 7. In this study the K Nearest Neighbor method is better in classifying majors on high Muhammadiyah 3 Jember school student data.

Keywords : Classification, Gaussian Naive Bayes , K-NN , Majoring students, SMOTE , Cross Fold Validation.