

Klasifikasi Performa Mahasiswa Berdasarkan Data Teks Forum Diskusi Online Menggunakan Multinomial Naive Bayes dengan Vektorisasi Pembobotan TF-IDF = Classification of Student Performance based on Text Data of Online Discussion Forums using Multinomial Naive Bayes with TF-IDF Weighting Vectorization

Rifqi Wazirsyah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529312&lokasi=lokal>

Abstrak

E-Learning Management System (EMAS) merupakan aplikasi yang dibuat oleh Universitas Indonesia dengan berbagai fitur salah satunya forum diskusi online. Dalam forum diskusi online, mahasiswa dapat membuat postingan-postingan dalam bentuk teks untuk bisa berdiskusi. Postingan-postingan dalam bentuk teks memiliki peran penting dalam meningkatkan performa mahasiswa yang terkhusus pada kelulusannya. Pada tugas akhir ini, Multinomial Naïve Bayes (MNB) digunakan untuk mengklasifikasi performa mahasiswa berdasarkan postingan-postingan dalam bentuk teks pada forum diskusi online. Sebelum dilakukan tahapan klasifikasi, postingan-postingan tersebut dilakukan preprocessing dan pemberian bobot kata pada teks menggunakan TF-IDF. Hasil TF-IDF dinyatakan dalam bentuk vektor-vektor, proses ini disebut dengan proses vektorisasi. Banyaknya dokumen dari data hasil vektorisasi TF-IDF yang digunakan yaitu sebanyak 228, dengan proporsi mahasiswa lulus dan tidak lulus secara berturut-turut, yaitu sebesar 219 dan 9. Pada data tersebut didominasi oleh mahasiswa lulus, artinya data tersebut tidak seimbang, sehingga diperlukan proses SMOTE untuk menyeimbangkan data. Kemudian, dilakukan implementasi model MNB pada 3 kasus pembagian data training dan data testing, yaitu 70%;30%, 80%:20% dan 90%:10%, dengan cara melatih model pada data training dan menguji model pada data testing untuk memperoleh klasifikasi performanya. Implementasi dilakukan sebanyak lima kali percobaan, sehingga didapatkan model MNB dapat mengklasifikasi performa mahasiswa dengan baik dan hasil kinerja model terbaik pada data testing 30% yaitu rata-rata akurasi sebesar 0,956, rata-rata recall sebesar 0,979, dan rata-rata f1-score sebesar 0,977. Namun rata-rata presisi terbaik didapatkan pada data testing 20%, yaitu sebesar 0,977.

.....E-Learning Management System (EMAS) is an application created by the University of Indonesia with various features, one of which is an online discussion forum. In online discussion forums, students can make posts in the form of text to be able to discuss. Posts in the form of text have an important role in improving student performance, especially at graduation. In this final project, Multinomial Naive Bayes (MNB) is used to classify student performance based on posts in text form on online discussion forums. Prior to the classification stage, the posts were preprocessed and assigned word weights to the text using TF-IDF. The results of TF-IDF are expressed in the form of vectors, this process is called the vectorization process. The number of documents from the TF-IDF vectorized data used is 228, with the proportion of students graduating and not graduating respectively, which is 219 and 9. SMOTE to balance data. Then, the implementation of the MNB model was carried out in 3 cases of distribution of training data and testing data, namely 70%; 30%, 80%:20% and 90%:10%, by training the model on the training data and testing the model on the testing data to obtain performance classification. The implementation was carried out five times, so that the MNB model was able to classify student performance well and the best model performance results were on 30% testing data, namely an average accuracy of 0.956, an average recall of

0.979, and an average f1-score of 0.956. 0.977. However, the best average precision was obtained at 20% testing data, which was 0.977.