

ABSTRAK

Mata kuliah merupakan unit pembelajaran di perguruan tinggi yang dipelajari oleh mahasiswa. Pada beberapa kasus mahasiswa, pemilihan mata kuliah seringkali berdasarkan pada faktor subjektif sehingga menyebabkan pemilihan mata kuliah menjadi kurang efektif. Penelitian ini bertujuan untuk memberikan alternatif solusi kepada mahasiswa dengan melakukan klusterisasi mata kuliah dengan fitur berupa dokumen referensi yang digunakan oleh dosen menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Metode ini termasuk dalam *unsupervised learning* dan menggunakan metode hierarkis dalam klusterisasi data. Dokumen referensi diperoleh dalam bentuk *e-book* dan dianalisis melalui tahapan *text preprocessing*, pembobotan TF-IDF, pengukuran kemiripan menggunakan *Cosine Similarity*, dan pengukuran jarak menggunakan *cosine distance*. Algoritma AHC kemudian diterapkan dengan tiga skenario *linkage*: *Single*, *Complete*, dan *Average*. Evaluasi dilakukan menggunakan *Silhouette Coefficient* untuk menilai kualitas hasil klusterisasi. Hasil dari tiga skenario pengujian terhadap 183 dokumen referensi mata kuliah dari 62 mata kuliah berbeda, didapatkan nilai *silhouette coefficient* terbaik yaitu, *average linkage* dengan nilai 0.265607. Kemudian pada skenario *complete linkage* 0.223531, dan *single linkage* 0,151134.

Kata kunci: Mata kuliah, *Text Mining*, *Clustering*, *TF-IDF*, *Cosine Similarity*, *Agglomerative Hierarchical Clustering*, *Silhouette Coefficient*, *single linkage*, *complete linkage*, *average linkage*.

ABSTRACT

Courses are learning units in higher education that are studied by students. In some cases, the selection of courses is often based on subjective factors, causing the selection of courses to be less effective. This study aims to provide an alternative solution to students by clustering courses with features in the form of reference documents used by lecturers using the Agglomerative Hierarchical Clustering (AHC) method. This method is included in unsupervised learning and uses a hierarchical method in data clustering. Reference documents are obtained in the form of e-books and analyzed through the stages of text preprocessing, TF-IDF weighting, similarity measurement using Cosine Similarity, and distance measurement using cosine distance. The AHC algorithm is then applied with three linkage scenarios: Single, Complete, and Average. Evaluation is carried out using the Silhouette Coefficient to assess the quality of the clustering results. The results of three test scenarios on 183 course reference documents from 62 different courses, obtained the best silhouette coefficient value, namely, the average linkage with a value of 0.265607. Then in the complete linkage scenario 0.223531, and single linkage 0.151134.

Keywords: *Courses, Text Mining, Clustering, TF-IDF, Cosine Similarity, Agglomerative Hierarchical Clustering, Silhouette Coefficient, single linkage, complete linkage, average linkage.*