

ABSTRAK

Penjualan berbasis online (*e-commerce*) kian banyak digunakan di bidang usaha. PT. Kanisius menggunakan aplikasi web untuk memudahkan para *customer* mereka untuk mencari, memilih dan membeli. Namun, web mereka belum dilengkapi dengan fitur rekomendasi kepada *customer*. Penelitian ini bertujuan untuk membangun sistem yang secara otomatis mampu mengelompokkan dan merekomendasikan buku berdasarkan sinopsis dan mengetahui tingkat akurasinya. Tahapan dimulai dari *preprocessing*, yaitu pemilihan data yang berdasarkan sinopsis buku, *tokenizing*, *stopword*, *stemming*, pembobotan kata, *principal component analysis*, normalisasi min-max, normalisasi *z-score* dan menghitung jarak antar data menggunakan *euclidean distance* dan *cosine similarity*. Proses selanjutnya adalah mengelompokkan data dengan menggunakan *agglomerative hierarchical clustering* yang memiliki 3 metode yaitu *single*, *average* dan *complete linkage*. Dilakukan 96 kali percobaan pengelompokan dan setiap percobaan dihitung nilai *sum of square error*.

Dari hasil percobaan tersebut, ditemukan hasil percobaan yang paling optimal pada percobaan ke-14 dengan *error* terkecil yaitu 3.0103. Percobaan tersebut menggunakan metode normalisasi min-max, penghitungan jarak menggunakan *euclidean distance* serta metode AHC *complete linkage*.

Kata kunci : Buku, *Tokenizing*, *Stopword*, *Stemming*, *Principal Component Analysis*, *Min-Max*, *Z-Score*, *Euclidean Distance*, *Cosine Similarity*, *Agglomerative Hierarchical Clustering*, *Sum of Square Error*

ABSTRACT

More businesses are using e-commerce nowadays. PT.Kanisius uses web to help the customers in searching, choosing, and buying online effortlessly. However, their web is not completed with recommendation feature yet. This research aims to build a system which can automatically cluster and present the book recommendation and figure out the accuracy using the agglomerative hierarchical clustering. The stages were started from preprocessing, tokenizing, stop word, stemming, word weighting, principal component analysis, normalization min-max, normalization score, and distance counting using Euclidean distance and cosine similarity. Data were collected by using the agglomerative hierarchical clustering which has 3 methods; they are single, average and complete linkage. 96 times of clustering trials were done and sum of square error value of each trial were counted.

From the trials, it was found that the 14th trial is the most optimum trial with the minimum error value 3.0103. The trial was presented by using the method of min-max normalization, distance counting based on euclidean distance, and AHC complete linkage.

Keywords : Book, Tokenizing, Stopword, Stemming, Principal Component Analysis, Min-Max, Z-Score, Euclidean Distance, Cosine Similarity, Agglomerative Hierarchical Clustering, Sum of Square Error