

INTISARI

PENGELOMPOKAN RESUME PELAMAR KERJA MENGGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING

Oleh

Ferdinand Pascanata Driyarkara

NIM 16/403678/PPA/05195

Dokumen resume berisi tentang informasi pelamar kerja dan digunakan sebagai salah satu acuan staf perekrut dalam melakukan perekrutan. Perekrutan dalam skala besar umumnya dilakukan oleh perusahaan penyedia tenaga kerja sebagai pihak ketiga antara perusahaan penyedia lowongan kerja dengan pelamarnya. Perusahaan tersebut menyimpan resume-resume pelamar yang diterima untuk berbagai lowongan kerja dari berbagai perusahaan. Ketika suatu lowongan kerja telah terpenuhi, resume yang berjumlah sangat banyak tersebut masih tersimpan sehingga dapat digunakan untuk lowongan lainnya. Namun, karena resume berjumlah sangat banyak dan terdiri dari berbagai jenis kemampuan kandidat, staf perekrut perlu mengelompokkannya terlebih dahulu berdasarkan kemiripannya. Pengelompokan ini dibutuhkan untuk memudahkan tahap awal perekrutan, yakni penyaringan kandidat.

Pengelompokan menggunakan *Agglomerative Hierarchical Clustering* (AHC) dilakukan terhadap beberapa konten penting pada resume. Dataset resume akan melalui tahap pra pemrosesan dan selanjutnya dikonversi ke bentuk vektor dengan *TF-IDF*. Penerapan AHC menggunakan perhitungan jarak *Euclidean Distance* dan penentuan kedekatan klasternya menggunakan *single linkage*, *complete linkage*, dan *average linkage*. Bentuk keluarannya adalah kluster hierarkis resume. Terdapat 100 dokumen resume yang digunakan sebagai sampel dan kemudian membentuk kelompok berbentuk hierarki.

Kualitas pengelompokan dihitung dengan metode validasi kluster partisi, *Silhouette Coefficient*, dan metode validasi kluster hierarkis, *Cophenetic Correlation Coefficient* (CPCC). Hasil pengelompokan resume dianggap kurang baik secara partisi karena walaupun memberikan nilai *silhouette index* yang cukup, kelompok-kelompok resume tidak seimbang. Namun, pengelompokan resume secara hierarki dapat digunakan karena uji validasi terhadap pohon hierarki pengelompokan resume memberikan nilai CPCC yang sangat baik.

Kata Kunci: Resume, Pengelompokan, AHC, Silhouette Coefficient, CPCC

ABSTRACT

GROUPING RESUME OF THE JOB APPLICANTS USING THE AGGLOMERATIVE HIERARCHICAL CLUSTERING METHOD

by

Ferdinand Pascanata Driyarkara

NIM 16/403678/PPA/05195

Resume document contains information about job applicants and used as a reference for recruiter staff in recruiting some employee candidates. Recruitment on a large scale is generally committed by employers providing employments as a third party between companies providing job vacancies and applicants. The company keeps applicants' resumes that they have received for various job openings from various companies. When a job vacancy has been fulfilled, all resumes that have been received are still stored so that it can be used for other vacancies. However, because there are a lot of resumes and consist of various types of candidate abilities, recruitment staffs need to group them in advance based on their similarity. This grouping is needed to facilitate the early stages of recruiting, i.e. candidate selection.

Grouping using the *Agglomerative Hierarchical Clustering (AHC)* method is used for some important content on the resume. The resume dataset will go through the pre-processing step and then be converted to vector with *TF-IDF*. *AHC* application uses *Euclidean Distance* calculation and determination of cluster proximity using *single linkage*, *complete linkage*, and *average linkage*. The output form is a hierarchical resume cluster. There are 100 resume documents used as samples and then form a hierarchical group.

The quality of the grouping is calculated by the validation method for the partitioned cluster, the *Silhouette Coefficient*, and the validation method for the hierarchical cluster, the *Cophenetic Correlation Coefficient (CPCC)*. The results of grouping resumes are considered unfavorable by partition because although they provide sufficient *silhouette index* values, resume groups are not balanced. However, hierarchical resume grouping can be used because the validation test of the resume grouping hierarchy tree gives an excellent *CPCC* value.

Keywords: Resume, Clustering, AHC, Silhouette Coefficient, CPCC