

Product Matching Book Analysis using CRISP-DM

12S45054 - Penambangan Data

Dibuat Oleh:

12S17017 - Yolanda Manurung

12S17027 - Stella Sitinjak

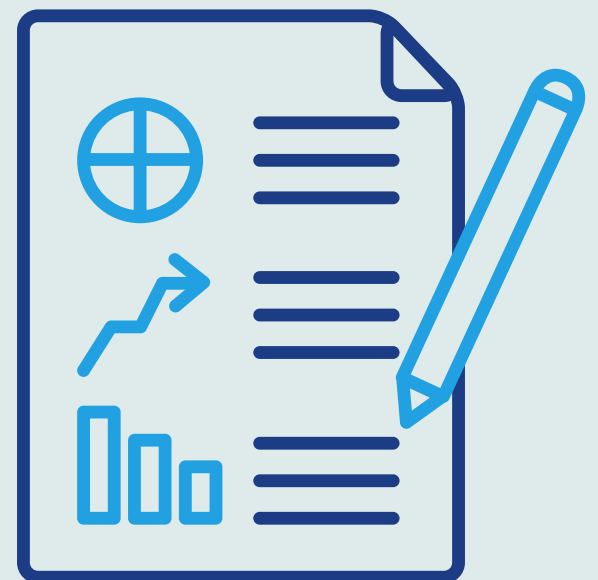
12S17059 - Ekis Naomi Lasma

Business Understanding

Data yang digunakan diambil dari Goodreads book review yang akan digunakan untuk membuat sebuah sistem *product matching* dengan menggunakan teknik *k-means clustering*.

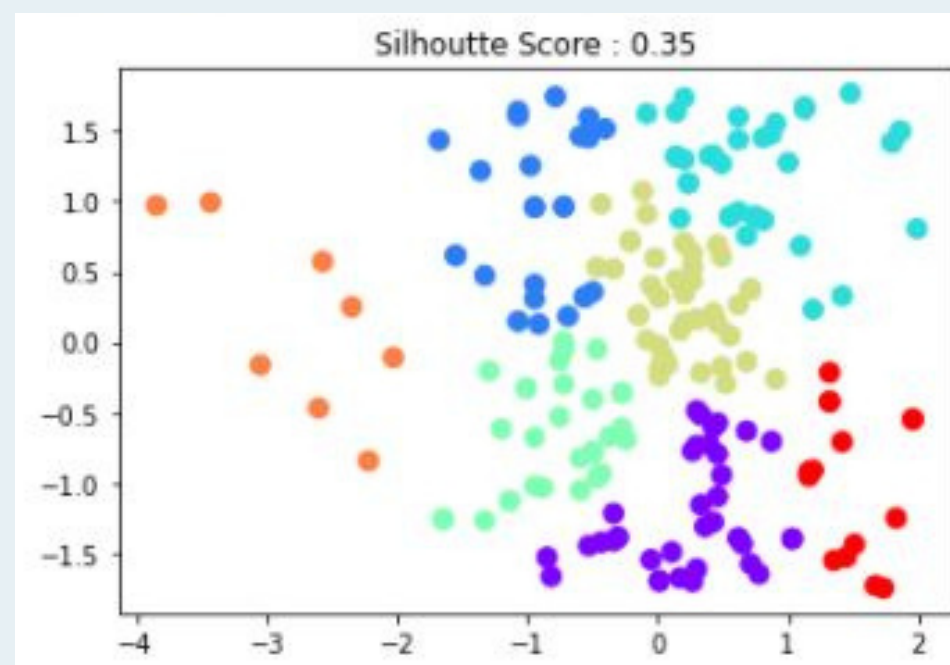
Data Understanding

Dataset tersebut terdiri dari 21,6k data dengan 15 atribut, 13 diantaranya bertipe data string dan 2 atribut lainnya bertipe data decimal.



Data Preparation

Pada dataset tersebut terdapat nilai *null* sehingga perlu dilakukan *cleaning* dan beberapa tahapan *preprocessing* yaitu *Remove Punctuation*, *Case Folding* dan *Stopwords Removing*.



Gambar 1. Hasil Cluster dengan algoritma agglomerative

Evaluation

Berdasarkan gambar 1 diketahui bahwa hasil *silhouette score* sama dengan 0.35 yang berarti model yang digunakan melakukan *clustering* dengan tepat

Data Modeling

Pembangunan model *product matching book analysis* menggunakan algoritma *agglomerative clustering* yang menghasilkan 7 cluster untuk melakukan *product matching* berdasarkan rating yang dimiliki tiap buku.

Deployment

Hasil dari penggunaan *k-means clustering* menghasilkan pengelompokan buku-buku berdasarkan review. Dokumentasi pengerjaan proyek meliputi dokumen akhir, dokumentasi *code* yang terdapat pada *repository* GitHub, poster, slide presentasi dan video presentasi.

Link GitHub :

<https://github.com/YolandaManurung/Product-Matching-Book-Analysis-using-K-means-Clustering>