

Product Clustering using K-Means Algorithm

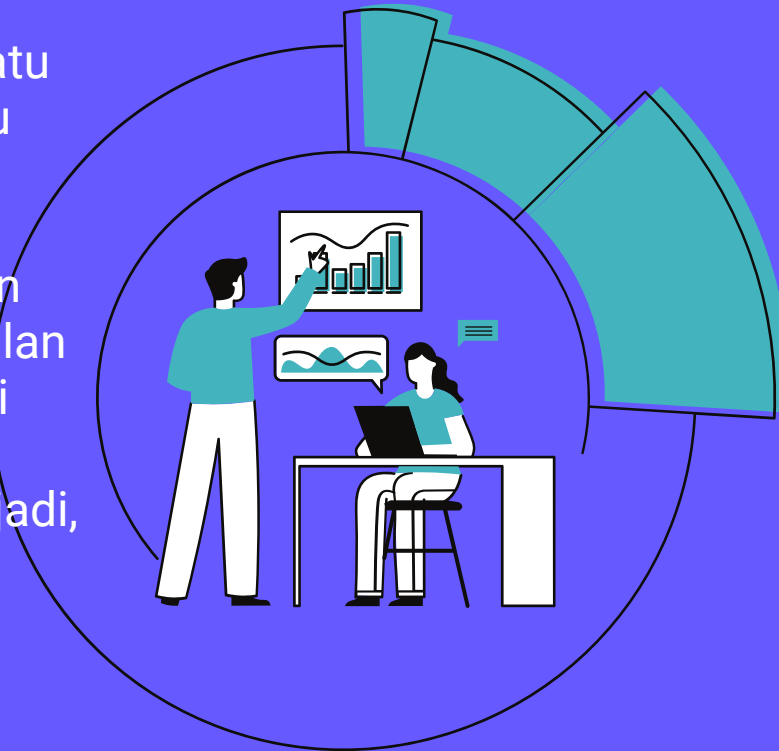
Oleh Kelompok 17



Institut Teknologi Del
12S4054 – Penambangan Data
Semester Gasal tahun 2020/2021

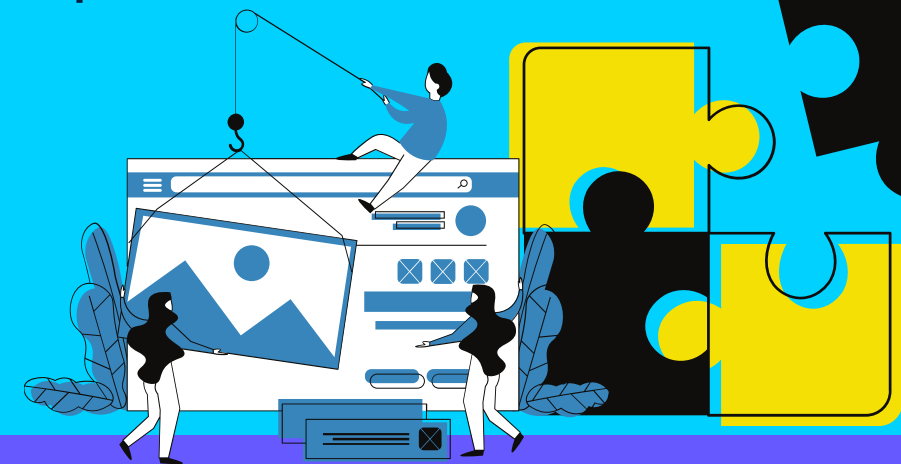
BUSINESS UNDERSTANDING

Dengan berbagai pilihan dan ketertarikan konsumen terhadap produk yang dibeli menjadi salah satu tantangan kepada para penjual atau pemilik toko dalam memilih produk mana yang sebaiknya disediakan dalam waktu tertentu sesuai dengan penjualan/minat konsumen. Penjualan produk terbanyak dapat kita ketahui dengan mengelompokkan produk-produk dari hasil transaksi yang terjadi, yang selanjutnya dapat digunakan sebagai dasar dalam mengambil keputusan.



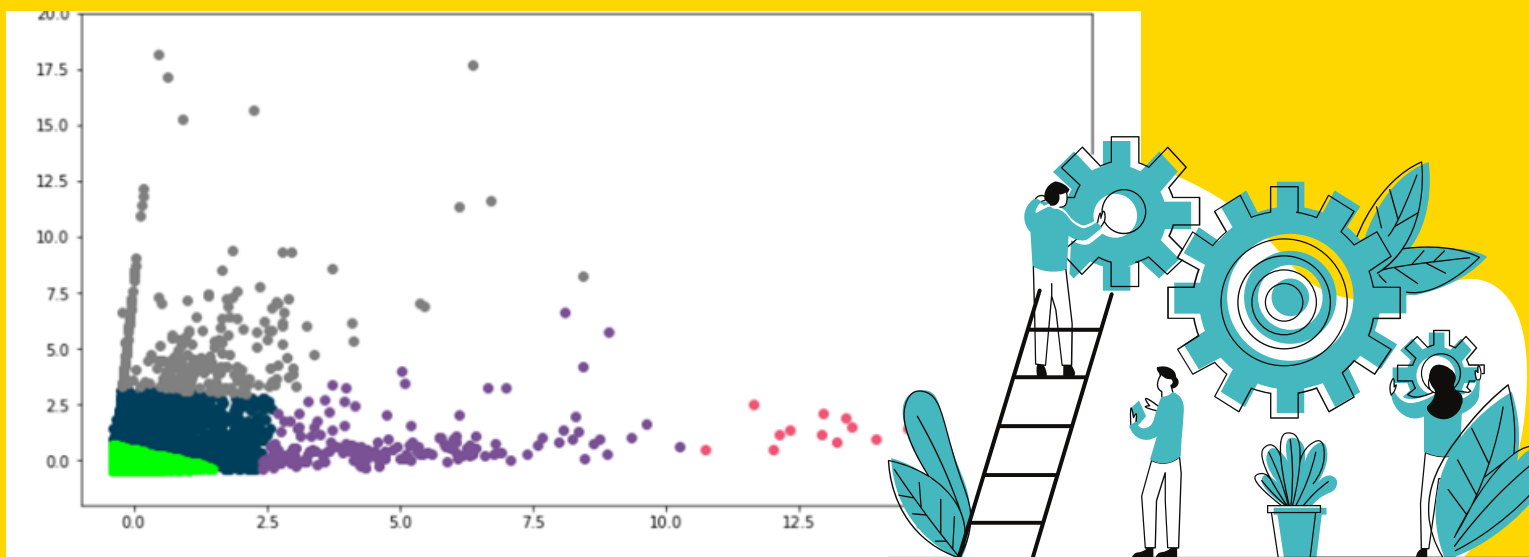
DATA UNDERSTANDING

Store Transaction Dataset adalah data transaksi penjualan sebuah toko yang menyimpan informasi terkait harga produk, jenis produk, jumlah produk terjual yang diambil dari website kaggle untuk mengidentifikasi Product Segments hasil transaksi penjualan produk dalam sebuah toko.



DATA MODELING

Menggunakan model Clustering dengan algoritma K-Means yang akan menghasilkan 5 Clustering untuk menentukan product clustering



DATA PREPARATION

data.sample(5)									
	STORECODE	QTY	VALUE	GRP	SGRP	SSGRP	CMP	MBRD	BRD
MONTH									
M1	P4	3	29	BISCUITS - CORE & NON CORE	CREAM	CREAM	MONDELEZ INTERNATIONAL	OREO	OREO VANILLA
M2	P3	6	1283	PACKAGED PURE GHEE	PACKAGED PURE GHEE	PACKAGED PURE GHEE	MOTHER DAIRY	AAREY	AAREY
M3	P4	4	40	DETERGENT CAKES/BARS	DETERGENT CAKES/BARS	DETERGENT CAKES/BARS	HARSH CLEAN DHAN PVT LTD	WOOSH	WOOSH
M3	P6	0	0	HAIR OILS PKTP(8/02)	RIGID	HAIR STYLING GEL CREAM	MARICO INDS	SET WET GELS	SET WET WET LOOK
M3	P6	21	252	SALTY SNACKS (2/97)	MIX/CHIVRA/CHANACHUR	MIX/CHIVRA/CHANACHUR	HALDIRAM	HALDIRAM	HALDIRAM PHALHARI CHIWDA

1. Cleaning yang dilakukan adalah dengan mengecek adanya nilai null, adanya data yang duplikat, dan mendeteksi adanya outlier di dalam dataset. Data Cleaning juga masih termasuk kedalam bagian dari Exploratory Data Analysis. Pada hasil exploratory data analysis yang dihasilkan bahwa tidak terdapat missing value pada dataset tersebut.

2. Feature selection dilakukan untuk membuang sebagian kolom atau atribut yang dianggap tidak terlalu dibutuhkan, dalam proses penambangan data, dengan menggunakan fungsi drop untuk men-drop beberapa atribut.

EVALUATION

Salah satu langkah yang dapat dilakukan dalam menentukan cluster yang optimal adalah dengan memplot grafik atau dikenal dengan elbow curve, dimana sumbu x mewakili jumlah cluster dan sumbu y mewakili matrik evaluasi (inersia).



DEPLOYMENT

Hasil Cluster dapat digunakan dan Menerapkan strategi yang tepat untuk mengambil keputusan supply produk dimasa mendatang, Pendokumentasian seperti Dokumen, Slide dan Poster



<https://github.com/PebriSinaga/Product-Clustering-Using-K-Means-Algorithm>