

## **PENERAPAN METODE K-MEANS UNTUK CLUSTERING HASIL PRODUKSI BERDASARKAN BERAT PRODUK**

Herman Purwoko Putro<sup>1)</sup>, Memed Saputra<sup>2)</sup>, Najmuddin<sup>3)</sup>

Universitas Ichsan Satya<sup>1)</sup>

Universitas Primagraha<sup>2,3)</sup>

hermanpurwoko1163@gmail.com<sup>1)</sup>, laboratorium890@gmail.com<sup>2)</sup>,

dede.najmuddin@gmail.com<sup>3)</sup>

### **ABSTRAK**

In the industrial world, especially in manufacturing companies that convert raw materials into finished goods. Quality is the main thing that absolutely must be fulfilled in all products. There are several factors or parameters that can be used as a basic benchmark in determining the suitability of a product. As a manufacturing company engaged in the motor vehicle tire manufacturing industry, of course there are many parameters used to determine product suitability. One of them is the weight of the product. This research aims to group products based on their weight, namely Normal, Overweight and Underweight. The data grouping process is carried out using Data Mining rules by utilizing Clustering techniques. The Data Mining process is carried out using Knowledge Discovery in Database (KDD) steps. The method used to obtain process results is by using the K-Means algorithm. Data processed using Microsoft Excel and implemented using Rapid Miner V7.4 software. The data used in this research was obtained from the Curing department with a data sample of 4758 records. The results of this research are to find out the percentage value of each product weight condition, how many products meet the product eligibility requirements and vice versa. From this, it is hoped that a conclusion can be drawn about how the production process is generally performed.

#### **Kata Kunci**

Data Mining, Clustering, Knowledge Discovery of Database (KDD), K-Means

## PENDAHULUAN

Analisis data adalah suatu kegiatan yang berkaitan dengan langsung dengan proses pengolahan data. Proses ini biasanya bertujuan untuk mencari suatu pola tertentu atau tren yang diinginkan dalam suatu database yang berukuran besar. Tujuan akhir dari kegiatan ini ditujukan untuk membantu pihak-pihak tertentu dalam pengambilan sebuah keputusan. Proses penggalian data-data yang tersedia menjadi sebuah informasi biasa disebut dengan Data Mining. Istilah ini mengacu kepada suatu proses penemuan pola dalam data yang sebelumnya tidak diketahui dan tidak ditemukan (Han et al., 2023). Proses penggalian data ini membutuhkan suatu perangkat dan metode tertentu untuk dapat mengolah data menjadi sebuah informasi yang dapat digunakan untuk menjadi sebuah alat bantu pengambilan keputusan. Semakin tinggi tingkat kompleksitas suatu informasi maka semakin rumit pula proses yang dilakukan.

Salah satu metode yang dapat digunakan dalam penggalian informasi ini adalah dengan menggunakan konsep clustering. Clustering secara umum dapat dikatakan sebagai proses pemisahan satu objek data kedalam beberapa bagian, dimana masing-masing bagian tersebut disebut dengan cluster. Dalam konsep cluster obyek-obyek yang berada dalam satu kelompok cenderung akan sama akan tetapi berbeda dengan objek-objek lain dalam cluster yang berbeda (Han et al., 2023). Clustering biasa disebut juga dengan segmentasi data, dalam aplikasinya clustering membagi set data yang besar kedalam beberapa kelompok sesuai kesamaan yang dimiliki.

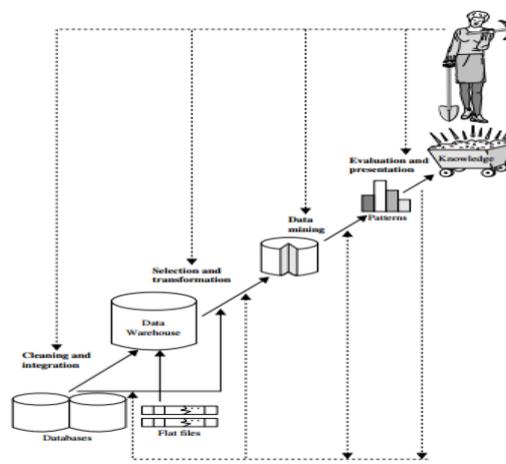
Tujuan dari penelitian ini adalah untuk menentukan seberapa besar kelayakan suatu produk untuk dilepas ke pasaran berdasarkan berat produk yang dihasilkan. Dalam penelitian ini berat produk tersebut dikelompokkan kedalam tiga kelompok yaitu “Dibawah Standar” (Underweight), “Standar” (Normal) dan “Diatas Standar” (Overweight). Dari hasil pengelompokan ini diharapkan dapat digunakan sebagai alat bantu dalam pengambilan keputusan untuk menekan hasil produk yang tidak memenuhi syarat kelayakan produk.

## TINJAUAN PUSTAKA

### Data Mining

Data mining merujuk pada metode analitika bisnis yang melampaui penghitungan, teknik deskriptif, pelaporan, dan metode berdasarkan aturan bisnis (Shmueli et al., 2018). Dalam istilah lain Data Mining dapat digambarkan sebagai proses penemuan suatu pola yang menarik atau proses penemuan suatu pengetahuan dari sebuah data yang berukuran sangat besar (Han et al., 2023; Septiana Ananda et al., 2023). Dalam istilah lain Data Mining merujuk kepada suatu istilah yang disebut Knowledge Discovery of Database (KDD) yang secara umum dapat diartikan sebagai proses penemuan suatu penemuan pada sebuah data (Haviluddin et al., 2021). Proses iterasi yang terdapat dalam Knowledge Discovery of Database (KDD) adalah sebagai berikut (Han et al., 2023):

1. Data Cleaning. Proses penghapusan data yang tidak konsisten atau data-data yang bersifat anomali.
2. Data Integration. Proses penggabungan data-data yang berasal dari banyak sumber data.
3. Data Selection. Proses menentukan data-data yang relevan yang akan digunakan dalam proses analisis data yang diambil dari database.
4. Data Transformation. Proses transformasi dan konsolidasi dalam bentuk yang sesuai untuk proses penggalian data melalui peringkasan atau dengan melakukan operasi agregasi.
5. Data Mining. Proses pengaplikasian sebuah metode atau algoritma untuk mengekstrak sebuah pola dari data yang tersedia.
6. Pattern Evaluation. Proses identifikasi pola-pola tertentu yang mewakili pengetahuan berdasarkan kondisi-kondisi atau ukuran-ukuran yang menarik.
7. Knowledge Presentation. Suatu teknik visualisasi dan representasi pengetahuan yang akan digunakan untuk mempresentasikan hasil-hasil penggalian pengetahuan.



Gambar 1 Knowledge Discovery of Database

### Clustering dan K-Means Clustering

Clustering dapat didefinisikan sebagai proses pengelompokan pola “X” menjadi berbeda segmen atau kelompok berdasarkan kedekatan atau kesamaan yang sesuai diantara pola-pola yang ada (Ye, 2014). Sebuah Clustering dianggap baik jika memiliki kesamaan yang tinggi dalam satu kelompok dan memiliki kesamaan pola yang rendah untuk dua atau lebih kelompok yang berbeda.

Tujuan dari Clustering adalah untuk meminimalkan jarak antar Cluster (Hermawati, 2013). Dalam istilah lain Clustering biasa disebut juga dengan Unsupervised Classification, yaitu suatu metode yang sedari awal tidak menentukan atau menggunakan informasi kelas dalam proses pengelompokannya. Salah satu metode dalam Clustering adalah K-Means Clustering, yaitu proses pengelompokan data melalui pendekatan partitional clustering berdasarkan sebuah titik pusat (centroid). Dalam implementasinya tiap data akan ditempatkan dalam Cluster dengan titik pusat (centroid) terdekat (Hermawati, 2013; Prasojo et al., 2019). Metode K-

Means menggunakan algoritma sebagai berikut (Haviluddin et al., 2021; Prasajo et al., 2019; Priyatman et al., 2019; Sarumaha, 2021):

- 1) Tentukan k sebagai jumlah Cluster atau jumlah kelompok yang ingin di bentuk.
- 2) Tentukan k Centroid (titik pusat cluster) awal. Penentuan centroid awal dilakukan secara random atau acak dari objek-objek yang tersedia sebanyak k cluster.
- 3) Hitung jarak setiap objek ke masing-masing Centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid, menggunakan rumus Euclidian Distance:

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Atau

$$d(x, y) = \left[ \sum_{i=1}^n |x_i - y_i|^2 \right]^{1/2}$$

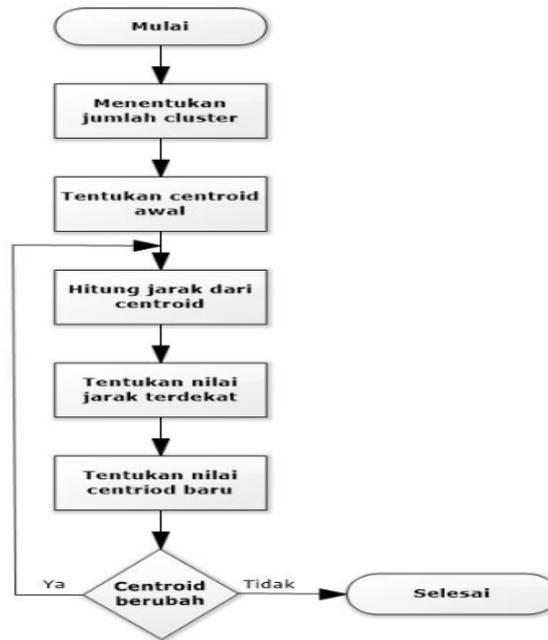
Nilai  $x_i$  : objek x ke-i  
 $y_i$  : data y ke-i n  
 $n$  : banyaknya objek

- 4) Alokasikan masing-masing objek ke dalam centroid yang paling terdekat. Untuk melakukan pengalokasian objek ke dalam masing-masing cluster pada saat iterasi secara umum dapat dilakukan dengan menyatakan setiap objek sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat cluster tersebut.
- 5) Melakukan iterasi kemudian menentukan posisi centroid baru dengan menggunakan persamaan :

$$v = \frac{\sum_{i=1}^n x_i}{n}$$

Nilai  $v$  : centroid pada cluster  
 $x_i$  : objek ke-i n  
 $n$  : banyaknya objek / jumlah objek yang menjadi anggota cluster

- 6) Mengulangi ke langkah 3 jika posisi centroid baru tidak konvergen. Pengecekan konvergensi dilakukan dengan membandingkan matriks group assignment pada iterasi sebelumnya dengan matrik group assignment pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma K-Means cluster dinyatakan sudah konvergen dan proses berhenti. Akan tetapi jika belum konvergen maka perlu proses pengulangan iterasi menuju kelangkah 3. Algoritma penentuan centroid dapat digambarkan sebagai berikut



Gambar 2 Algoritma K-Means

**Kerangka Pemikiran**

Kerangka penelitian adalah diagram yang menjelaskan secara umum alur perjalanan sebuah penelitian. Kerangka pemikiran dibuat berdasarkan pertanyaan penelitian dan merepresentasikan suatu himpunan dari beberapa konsep serta hubungan dari konsep tersebut. Dalam penelitian ini kerangka pemikiran yang digunakan adalah sebagai berikut.

Permasalahan		
Mengetahui banyaknya produk yang memenuhi standar kelayakan produksi maupun tidak.		
Tujuan		
Megelompokan produk berdasarkan berat yang dimiliki		
Eksperimen		
Sumber Data	Metode	Implementasi
Sumber data produk dan berat produk	K-Means	Rapid Miner
Hasil		
Informasi banyaknya produk berdasarkan tiga kriteria berat produk (O, N, U)		
Manfaat		
Dapat digunakan sebagai alat bantu manajemen dalam menekan hasil produksi yang tidak sesuai dengan standar kelayakan produksi.		

Gambar 3. Kerangka Pemikiran.

**METODE PENELITIAN**

Dalam penelitian ini interface yang digunakan untuk menguji proses data clustering dan mempresentasikan visualisasi dari hasil pengujian adalah dengan menggunakan aplikasi Rapid Miner. Rapid Miner merupakan platform perangkat lunak yang menyediakan sebuah sistem terintegrasi untuk proses-proses machine learning, penggalian data dan analisis prediksi. Aplikasi ini dapat digunakan untuk riset, pendidikan, pelatihan, prototipe dan pengembangan aplikasi yang mendukung proses termasuk didalamnya preparasi data, visualisasi, validasi dan optimalisasi (Savaram, 2023; Septiana Ananda et al., 2023)

**HASIL DAN PEMBAHASAN**

Peneliti melakukan beberapa prosedur dalam mengimplementasikan pengujian data clustering, secara umum prosedur tersebut mengacu kepada Knowledge Discovery of Database (KDD) dimana prosedur diawali dari penyiapan data hingga presentasi visual dari pengujian yang dilakukan. Jumlah data yang digunakan dalam penelitian ini yaitu 4758 record yang berarti sama dengan 4758 produk barang jadi yang diproduksi. Data didapat dari hasil penimbangan yang di extract langsung dari mesin dalam bentuk file Excel berekstensi .Xls. Sebelum didapatkan hasil penelitian perlu dilakukan beberapa prosedur pengolahan data. Prosedur-prosedur tersebut adalah (Darwis et al., 2021; Sarumaha, 2021) :

**A. Persiapan Data Produk**

Dataset yang digunakan dalam pengujian ini adalah dataset yang didapatkan dari departemen Final Inspection, dataset tersebut berisi beberapa attribut diantaranya Barcode, tanggal proses, kode produk, berat produk, standar berat yang di sarankan, batas berat minimum dan batas berat maksimum. Pengujian ini ditujukan untuk mengetahui seberapa besar pengelompokan produk-produk tersebut berdasarkan berat yang dimiliki.

BARCODE	EVENTDATE	MACHINEID	WEIGHT_ACTUAL	WEIGHT_REF	WEIGHT_MIN	WEIGHT_MAX	STATUS	FLAG	SHIFT	RESEND	ITEM_CODE	MONTH_YEAR	FLAG_STS
AA0011729519	01/02/2017 05.53	172.16.180.36	13,4725	13,34941509	12,81543827	13,88339138	0	0	3	0	AA001	201702	Y
B285 1707124	01/02/2017 05.53	172.16.180.36	9,675167	9,599382914	9,215407372	9,983358383	0	0	3	0	B285	201702	Y
XXXXX2017215564	01/02/2017 05.56	172.16.180.36	14,81217	0	0	0	3	0	3	0	XXXXX	201702	N
XXXXX2017215567	01/02/2017 05.56	172.16.180.36	14,81217	0	0	0	3	0	3	0	XXXXX	201702	N
BM1611712039	01/02/2017 05.56	172.16.180.36	13,0365	12,8813633	12,36610889	13,39661789	0	0	3	0	BM161	201702	Y
CA0081703413	01/02/2017 05.57	172.16.180.36	11,10367	11,04391929	10,60216236	11,48567581	0	0	3	0	CA008	201702	Y

Gambar 4. Dataset Awal

**B. Data Cleansing dan Preprocessing**

Setelah dataset tersedia proses berikutnya adalah proses data cleansing, yaitu proses pemisahan data-data yang berpotensi menyebabkan kesalahan pada proses pengujian. Kesalahan dalam salah satu nilai atau attribute atau informasi kelas dikenal dengan istilah noise.

Data-data yang bersifat noise tersebut dapat dibenahi dengan beberapa cara, seperti:

- Mengabaikan nilai-nilai yang hilang.
- Memprediksi nilai yang hilang berdasarkan nilai-nilai yang diketahui.

- Memperlakukan data yang hilang sebagai sebuah nilai khusus.
- Menghitung nilai rata-rata dengan menggunakan metode Bayesian.
- Menghilangkan nilai-nilai yang bersifat noise.

Dalam penelitian ini metode yang digunakan untuk meminimalisir timbulnya kesalahan karena adanya noise adalah dengan menghilangkan data-data yang bersifat noise tersebut dari dataset.

Pilihan ini diambil karena data-data yang bersifat noise tersebut tidak bisa di prediksi nilainya atau diperlakukan sebagai sebuah nilai khusus. Selain itu pilihan ini diambil sekaligus untuk menghindari adanya kesalahan data terutama untuk data-data hasil proses produksi.

BARCODE	EVENTDATE	MACHINEID	WEIGHT_ACTUAL	WEIGHT_REF	WEIGHT_MIN	WEIGHT_MAX	STATUS	FLAG	SHIFT	RESEND	ITEM_CODE	MONTH_YEAR	FLAG_STS
AA0011729519	01/02/2017 05.53	172.16.180.36	13,4725	13,34941509	12,81543827	13,88339138	0	0	3	0	AA001	201702	Y
B285 1707124	01/02/2017 05.53	172.16.180.36	9,675167	9,599382914	9,215407372	9,983358383	0	0	3	0	B285	201702	Y
XXXXX2017215564	01/02/2017 05.56	172.16.180.36	14,81217	0	0	0	3	0	3	0	XXXXX	201702	N
XXXXX2017215567	01/02/2017 05.56	172.16.180.36	14,81217	0	0	0	3	0	3	0	XXXXX	201702	N
BM1611712039	01/02/2017 05.56	172.16.180.36	13,0365	12,8813633	12,36610889	13,39661789	0	0	3	0	BM161	201702	Y
CA0081703413	01/02/2017 05.57	172.16.180.36	11,10367	11,04391929	10,60216236	11,48567581	0	0	3	0	CA008	201702	Y

Gambar 1. Penghapusan noise dari dataset

Sehingga dataset yang didapatkan adalah sebagai berikut :

BARCODE	EVENTDATE	MACHINEID	WEIGHT_ACTUAL	WEIGHT_REF	WEIGHT_MIN	WEIGHT_MAX	STATUS	FLAG	SHIFT	RESEND	ITEM_CODE	MONTH_YEAR	FLAG_STS
AA0011729519	01/02/2017 05.53	172.16.180.36	13,4725	13,34941509	12,81543827	13,88339138	0	0	3	0	AA001	201702	Y
B285 1707124	01/02/2017 05.53	172.16.180.36	9,675167	9,599382914	9,215407372	9,983358383	0	0	3	0	B285	201702	Y
BM1611712039	01/02/2017 05.56	172.16.180.36	13,0365	12,8813633	12,36610889	13,39661789	0	0	3	0	BM161	201702	Y
CA0081703413	01/02/2017 05.57	172.16.180.36	11,10367	11,04391929	10,60216236	11,48567581	0	0	3	0	CA008	201702	Y

Gambar 2. Dataset tanpa noise

Setelah tahapan diatas dilakukan proses berikutnya adalah preprocessing yaitu tahapan yang dilakukan untuk menganalisis data apa saja yang sebenarnya dibutuhkan dalam proses K-Means(Nurjanto, 2013).

Dari dataset diatas, sesuai dengan tujuan pengujian maka atribut-attribut yang digunakan adalah :

BARCODE	MACHINEID	WEIGHT_ACTUAL	WEIGHT_MIN	WEIGHT_MAX	ITEM_CODE
AA0011729519	172.16.180.36	13,4725	12,81543827	13,88339138	AA001
B285 1707124	172.16.180.36	9,675167	9,215407372	9,983358383	B285
BM1611712039	172.16.180.36	13,0365	12,36610889	13,39661789	BM161
CA0081703413	172.16.180.36	11,10367	10,60216236	11,48567581	CA008

Gambar 3. Dataset yang siap digunakan

### C. Transformation

Proses *transformation* pada metode *K-Means* harus memenuhi kaidah *Euclidian Distance*, sehingga parameter-parameter nilai yang akan digunakan dalam pengujian dan tidak bertipe *numeric* haruslah dirubah terlebih dahulu menjadi nilai bertipe *numeric* agar dapat diimplementasikan dalam persamaan *Euclidian Distance*.

Dataset untuk attribute berat akan dikelompokkan kedalam 3 (tiga) kategori yaitu:

1. Overweight (O)  
 Kondisi  $WEIGHT\_ACTUAL > WEIGHT\_MAX$
2. Normal (N)  
 Kondisi  $WEIGHT\_ACTUAL \geq WEIGHT\_MIN$  dan  $WEIGHT\_ACTUAL \leq WEIGHT\_MAX$
3. Underweight (U)  
 Kondisi  $WEIGHT\_ACTUAL < WEIGHT\_MIN$

Berikutnya adalah penentuan inisiasi dari beberapa attribut yaitu untuk attribut MACHINEID dan ITEMCODE. Proses inisiasi dilakukan secara bebas tanpa ada kriteria tertentu. Proses inisiasi yang dilakukan yaitu :

a. MACHINE\_ID

Hasil inisiasi untuk MACHINE\_ID. ID diurutkan berdasarkan nilai MachineID.

MachineID	ID
172.16.180.28	1
172.16.180.29	2
172.16.180.30	3
172.16.180.36	4

Gambar 4. Inisiasi MACHINE\_ID

b. ITEM\_CODE

Hasil inisiasi untuk ITEM\_CODE. ID diurutkan berdasarkan ItemCode.

ItemCode	ID
A096	1
A105	2
A252	3
A253	4
A291	5
A556	6
A584	7

Gambar 5. Inisiasi ITEM\_CODE

c. WEIGHT\_ACTUAL

Untuk attribute WEIGHT\_ACTUAL perlu dilakukan logika pengelompokan sebagai berikut :

Jika

$WEIGHT\_ACTUAL \geq WEIGHT\_MIN; WEIGHT\_ACTUAL \leq WEIGHT$

Maka nilai = 'N'

Jika  $WEIGHT\_ACTUAL < WEIGHT\_MIN$

Maka nilai = 'U'

Jika  $WEIGHT\_ACTUAL > WEIGHT\_MAX$

Maka nilai = 'O'

Dan dari hasil proses diatas didapatkan sebuah dataset baru.

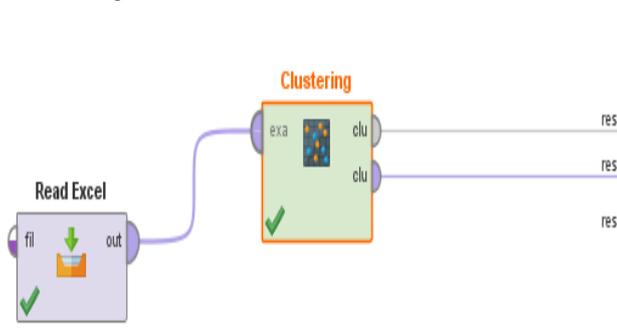
BARCODE	MACHINE_ID	WEIGHT_ACTUAL	ITEM_CODE	RefWeight
BM1761712484	6	9,7655	58	N
BM0491707398	6	9,6718	43	N
BM1761712485	6	9,758099	58	N
B256 1714884	6	9,1427	32	N
BM0491707399	6	9,7463	43	N
BM1761712486	6	9,7711	58	N
B256 1714885	6	9,1167	32	N

Gambar 6. Dataset yang telah ditransformasi

Dataset inilah yang nantinya siap diimplementasikan pada *interface* Rapid Miner

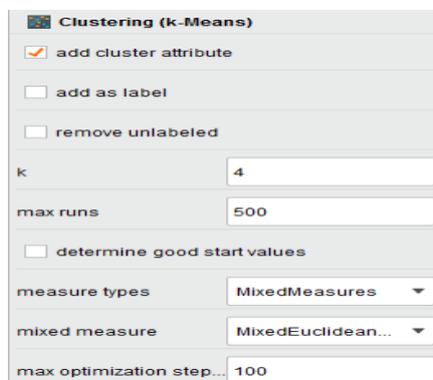
**D. Data Mining dan Presentation**

Dataset diimplementasikan pada perangkat lunak Rapid Miner. Dengan menggunakan pendekatan *K-Means Clustering*



Gambar 7. Implementasi di Rapid Miner

Parameter pengujian yang digunakan



Gambar 8. Parameter pengujian

Hasil Pengujian dari dataset pada perangkat lunak Rapid Miner untuk pemodelan *cluster*, nilai *centroid* yang digunakan dan grafik penyebaran data untuk perbandingan **RefWeight** terhadap **Cluster** adalah sebagai berikut:

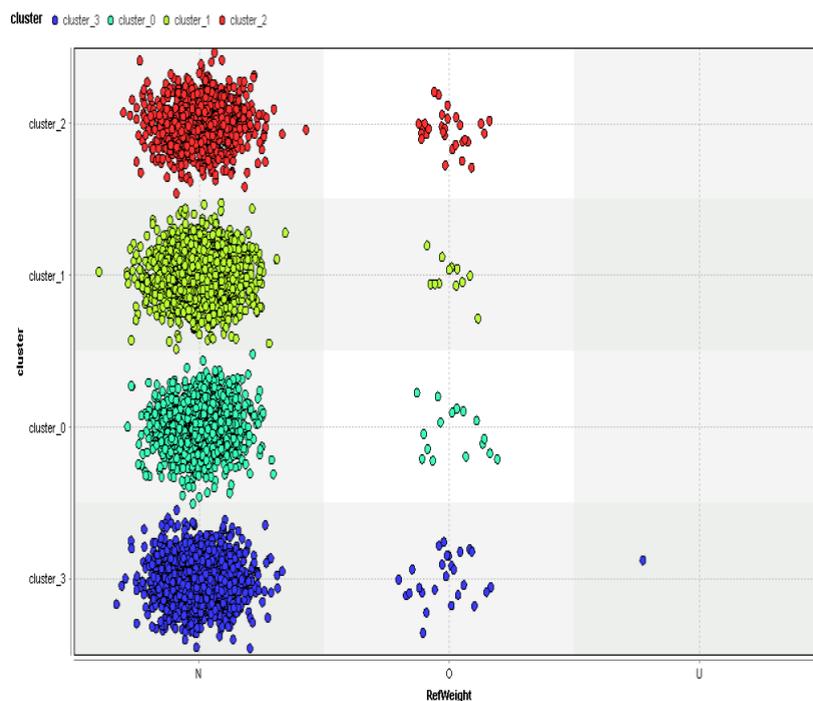
### Cluster Model

```
Cluster 0: 1053 items
Cluster 1: 1606 items
Cluster 2: 842 items
Cluster 3: 1257 items
Total number of items: 4758
```

Gambar 9. Hasil Clustering

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
BARCODE	2300.555	2422.409	2393.110	2353.368
MACHINE_ID	4.397	3.555	4.228	4.553
WEIGHT_ACTUAL	10.829	9.908	13.196	10.067
ITEM_CODE	37.399	90.697	11.127	63.973
RefWeight	0.015	0.007	0.034	0.022
ID_WEIGHT	1.015	1.007	1.034	1.022

Gambar 10. Nilai Centroid Dataset



Gambar 11. Grafik perbandingan RefWeight dan Cluster

Dari hasil clustering terhadap contoh data sebanyak 4758 *record* untuk melihat pengelompokan produk berdasarkan tiga kriteria (*Overweight*, Normal, *Underweight*) didapatkan nilai seperti yang terdapat pada gambar 16 berikut.

Index	Nominal value	Absolute count
1	N	4674
2	O	83
3	U	1

Gambar 12. Hasil pengelompokan berat produk

Dari tabel diketahui bahwa jumlah produk untuk kategori N (Normal) adalah 4674 atau 98,23%, kategori O (*Overweight*) adalah 83 atau 1,74 % dan untuk kategori U (*Underweight*) adalah 1 atau 0,03 %.

Hasil ini dapat menjelaskan tujuan penelitian, bahwasanya produk-produk yang dihasilkan jika dinilai dari parameter berat produk maka secara umum telah sesuai dengan standar yang berlaku. Berat produk yang menyimpang dari standar yang berlaku tetap ada walaupun nilainya kecil.

## KESIMPULAN

Setelah dilakukan proses transformasi dan pengelompokan serta pengujian terhadap data produk dengan membagi data mejadi 4 cluster untuk menentukan berat produk maka didapatkan hasil 98,23% untuk kategori N (Normal), 1,74 % untuk kategori kategori O (*Overweight*) dan untuk kategori U (*Underweight*) adalah 0,03 %. Dari hasil ini dapat dikatakan jika secara umum pengkategorian produk berdasarkan berat yang dihasilkan menunjukkan bahwa sebagian besar produk telah memenuhi salah satu kelayakan standar produk yaitu dari sisi berat. Proses analisis lebih lanjut tetap harus dilakukan oleh pihak perusahaan agar nilai penyimpangan standar dapat lebih ditekan lagi.

## REFERENSI

- Darwis, M., Hasibuan, L. H., Firmansyah, M., & Ahady, N. (2021). Implementation of K-Means Clustering Algorithm in Mapping the Groups of Graduated or Dropped-out Students in the Management Department of the National University. *JISA (Jurnal Informatika Dan Sains)*, 04(01), 1–9.
- Han, J., Pei, J., & Tong, H. (2023). *Data Mining Concepts and Techniques Fourth Edition*. Massachusetts:Elsevier.
- Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir.

*Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 16(1).  
<https://doi.org/10.30872/jim.v16i1.5182>

Hermawati, F. A. (2013). *Data Mining*. Penerbit Andi.

Nurjanto, F. D. (2013). *Tahap-tahap K-Means Clustering*.

Diakses tanggal 10 Januari 2024 dari <https://fadlikadn.wordpress.com/2013/06/14/tahap-tahap-k-means-clustering/>

Prasojo, R., Utami, Y. R. W., & Vlandari, R. T. (2019). Implementasi K-Means Clustering Pada Pengelompokan Potensi Kerjasama Pelanggan. *Jurnal TIKomSIN*, 7.

Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1). <https://doi.org/10.26418/jp.v5i1.29611>

Sarumaha, N. I. (2021). Implementasi Algoritma K-Means Clustering Pada Analisa Impor Beras. *JUSSI: Jurnal Sains Dan Teknologi Informasi*, 1(1), 19–27.

Savaram, R. (2023). *RapidMiner Tutorial / What Is RapidMiner - A Complete Guide - 2024*. Diakses tanggal 10 Januari 2024 dari <https://mindmajix.com/rapidminer-tutorial>

Septiana Ananda, P., Sediono, E., & Sembiring, I. (2023). KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal. *Jurnal BATIRSI*, 6(2), 8–13.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Kenneth C. Lichtendahl, J. (2018). *DATA MINING FOR BUSINESS ANALYTICS Concepts, Techniques, and Applications in R* (First Edition). New Jersey: John Wiley & Sons, Inc.

Ye, N. (2014). *Data Mining: Theories, Algorithms, and Examples*. Florida: CRC Press.