

## **Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma *K-means***

**Dodi Nofri Yoliadi<sup>1</sup>**

Fakultas Ushuluddin Adab dan Dakwah, UIN Mahmud Yunus Batusangkar

<sup>1</sup> dodinofriyadi@iainbatusangkar.ac.id

Diterima: 28 Feb 2023 | Direvisi: 28 Feb 2023

Disetujui: 28 Feb 2023 | Dipublikasi: 28 Feb 2023

### **Abstrak**

Pada penelitian ini dilakukan pengumpulan dan pengolahan data penjualan barang elektronik dengan CV. Berkah Elektronik, data yang diperoleh kemudian diintegrasikan dengan algoritma *K-means* dalam mendapatkan pengetahuan tentang produk elektronik mana yang sangat laku dipasaran dengan baik dan yang tidak. Pada penelitian ini digunakan metode *clustering* dengan software Tanagra 1.4.48, dan algoritma *K-means* digunakan sebagai algoritma untuk menarik kesimpulan barang mana yang sangat laku dan mana yang tidak di input adalah harga produk, barang dan penjualan barang. Dan dari hasil pengujian dan pengujian manual dengan aplikasi Tanagra, dihasilkan *cluster* yang sama, yaitu produk yang tidak laku sebagai *Cluster\_KMeans\_1* dan produk yang laku sebagai *Cluster\_KMeans\_2*.

**Kata kunci:** *K-means*, *Clustering*, *Penjualan*, *Data Mining*, *Tanagra*

### **Abstract**

*In this study, data collection and processing were performed on sales of CV electronic products. Belka Electronics. The data obtained is aggregated using the K-means algorithm to understand which electronic products are selling well and which are not. In this study, we used a clustering method with Tanagra 1.4.48 software and used the K-means algorithm as the algorithm to conclude which items sold well and which did not, i.e., the commodity price. Sale of goods or merchandise. And from the test results using the Tanagra application and the manual tests, the same clusters are created: unsold products*

*(Cluster\_KMeans\_1) and selling products (Cluster\_KMeans\_2).*

**Keywords:** *K-means*, *Clustering*, *Sales*, *Data Mining*, *Tanagra*

### **I. PENDAHULUAN**

*Data mining* merupakan sesuatu teknik yang digunakan untuk menemukan informasi dalam suatu basis data. *Data mining* dalam prakteknya mengimplementasikan berbagai teknik diantaranya; Kecerdasan buatan, Pembelajaran mesin/ machine learning, Kecerdasan Buatan dan teknik statistik yang digunakan untuk mengekstraksi data dan mengidentifikasi data untuk kebutuhan informasi dari sebuah data base yang besar [1].

*Data mining* ialah proses mengekstraksi data dari sekumpulan data/informasi yang begitu besar dan tak terbatas. Pada dasarnya *data mining* adalah langkah KDD (*Knowledge Discovery In Databases*). Layanan informasi merupakan kegiatan dari beberapa proses diantaranya; pembersihan data, integrasi, pemilihan data, transformasi, data mining, penilaian pola dan penyajian data [2].

Saat ini sistem terkomputasi sudah banyak digunakan pada berbagai perusahaan bisnis guna mendapatkan informasi-informasi penting mengenai kegiatan pemasaran. Salah satu perusahaan yang menerapkan sistem komputasi ini ialah bisnis yang berada pada bidang penjualan produk barang elektronik. Fungsi dari pengolahan data bagi perusahaan tersebut adalah untuk mendapatkan informasi yang tepat, akurat serta mudah dimengerti untuk melihat penjualan barang elektronik. Untuk mendapatkan informasi tersebut,

digunakan berbagai teknik yang salah satunya ialah teknik *clustering* yang tersedia di data mining, yaitu algoritma *K-means clustering*.

Algoritma *K-means* merupakan salah satu bentuk teknik *clustering* yang berkembang pesat, hal tersebut terjadi karena teknik tersebut sangat mudah untuk di implementasikan dan digunakan. Pada teknik tersebut objek data yang ada dikelompokkan dalam bentuk *cluster-cluster*. Untuk penggabungan, nilai *k* harus dilakukan *clustering* terlebih dahulu. *K-means* ialah metode *clustering* yang mudah dan sederhana serta gampang untuk diimplementasikan karena kelebihanannya ialah bisa melakukan klasifikasi data dalam jumlah yang banyak/ besar dan membuat *cluster* dengan cepat, tepat dan akurat. Tujuan dari algoritma *K-means* adalah melakukan *clustering* objek sedemikian rupa hingga jarak setiap objek ke pusat *cluster* minimal. Pengelompokkan data dapat diaplikasikan untuk menganalisa tingkat penjualan barang elektronik di CV. Berkah Elektronik. Semakin cepat diketahui tingkat penjualan barang elektronik semakin baik karena pihak perusahaan dapat melakukan perbaikan. Menganalisis tingkat penjualan produk yang tersedia dengan melakukan teknik algoritma *K-means clustering* diharapkan dapat memberikan informasi yang akurat dan tepat tentang jenis barang elektronik yang paling banyak terjual dan dibeli serta diminati oleh pembeli (*customer*).

## II. TINJAUAN PUSTAKA

### A. Data mining

*Data mining* ialah sesuatu teknik yang dipergunakan dalam menemukan informasi dalam suatu *database*. *Data mining* dalam prakteknya mengimplementasikan berbagai teknik diantaranya; Kecerdasan buatan, Pembelajaran mesin/ machine learning, Kecerdasan Buatan dan teknik statistik yang digunakan untuk mengekstraksi data dan mengidentifikasi data untuk kebutuhan informasi dari sebuah data base yang besar [1].

*Data mining* pada prinsipnya merupakan proses pengalihan data/ informasi dari database yang tak terbatas dan besar. Faktanya, *Data mining* ialah langkah menuju *Knowledge Discovery In Databases* (KDD). Layanan informasi berbagai proses antara lain; pembersihan data, integrasi, pemilihan data, transformasi, penambahan data, penilaian pola dan penyajian data [7].

### B. Data mining Siklus

Menurut CRISP-DM, siklus hidup *data mining* dibagi menjadi 6 tahap. Setiap tahapan saling ketergantungan antara satu tahapan dengan tahapan-tahapan selanjutnya. Output tahapan sebelumnya akan mempengaruhi secara tidak langsung tahapan yang akan terjadi berikutnya.

6 dalam tahapan CRISP- DM yaitu [1]:

- 1) Step pada Fase Pemahaman Bisnis
- 2) Step pada Fase Pemahaman Data
- 3) Step pada Fase Pengolahan Data
- 4) Step pada Fase Pemodelan
- 5) Step pada Fase Evaluasi
- 6) Step pada Fase Penyebaran

### C. Clustering

*Clustering* merupakan teknik yang populer dan banyak digunakan dalam proses data mining. Berbagai usaha telah dilakukan oleh Para peneliti di bidang *data mining* untuk menyempurnakan model klaster, karena metode pengembangannya masih bersifat heuristik. Perhitungan jumlah klaster optimal dan klaster terbaik masih dalam proses. Oleh karena itu, dengan metode yang diterapkan sekarang, tidak dapat memastikan bahwa hasil *clustering* yang didapatkan akan optimal. Namun, dari segi praktis, hasil yang dicapai biasanya cukup baik [2].

Tujuan yang ditargetkan dari *clustering* ialah untuk mengklasifikasikan data/ objek kedalam *cluster-cluster* atau kelompok-kelompok sedemikian rupa hingga setiap kelompok berisi data dengan karakteristik yang identik. Pada tahapan proses *clustering* kita mengupayakan meletakkan objek yang memiliki ciri khas yang identik dalam *cluster* yang sama kemudian membuat jarak antara satu *cluster* dengan *cluster* lainnya semakin luas [2].

### D. Cluster Analisis

Pada penelitian ini menggunakan metode *clustering*, pada dasarnya metode *clustering* melakukan identifikasi terhadap objek dengan karakteristik umum tertentu, setelah itu karakteristik tersebut digunakan sebagai *eigenvector* atau *centroid* [6]. *Cluster Analisis* digunakan untuk memunculkan kumpulan objek yang serupa/ identik sehingga objek dalam satu *cluster* serupa (atau terkait) dengan yang lain dan objek dengan *cluster* berbeda akan terpisah dengan *cluster* yang sesuai [3].

Tujuan proses *cluster* analisis ialah langkah/ cara yang digunakan untuk mengatur jarak antar *cluster* baik untuk meminimalkan jarak ataupun memaksimalkan jarak antar *cluster* yang ada. Proses *Clustering* bisa dikatakan bersifat kuantitatif (numerik), kualitatif (kategorik), atau gabungan keduanya. Data yang diolah bisa berasal dari hasil observasi dari sebuah proses yang terjadi. Setiap pengamatan dapat memiliki  $n$  variabel pengukuran dan vektor  $n$ -dimensi yang dikelompokkan  $Z_k = [Z_{1k}, \dots, Z_{nk}]^T, Z_k \in \mathbb{R}^n$ . Himpunan pengamatan  $N$  diberikan oleh  $Z = \{Z_k \mid \text{menunjukkan } k = 1, 2, \dots, N\}$  dan direpresentasikan sebagai matriks  $n \times N$  [1].

### E. Algoritma *K-means*

Algoritma *K-means* adalah algoritma yang paling banyak digunakan dan diimplementasikan pada dunia industri modern. Dalam implementasiannya Algoritma *K-means* berdasarkan pada prinsip yang sederhana dan mudah untuk dimengerti. Pada langkah awal cukup dengan menentukan jumlah *cluster* yang digunakan, dari setiap objek/ elemen pertama dari sebuah *cluster* dapat dijadikan pusat dari *cluster* tersebut [10].

Beberapa alternatif untuk menerapkan *K-means* telah diusulkan, dengan beberapa teori evolusi komputasi terkait. Ini termasuk pilihan [12]:

- Distance space* dalam mengkalkulasi jarak data terhadap *centroid*
- Metode untuk memetakan data yang masuk ke masing-masing *cluster*
- Fungsi dari tujuan gunakan.

*K-means* dikategorikan ke dalam pengelompokan partisi/ *cluster*. Data-data yang tersedia harus dikelompokkan ke dalam *cluster-cluster* tertentu dan semua data milik *cluster* tertentu dapat dipindahkan ke *cluster* lain dalam satu langkah proses, di langkah berikutnya. *K-means* mengelompokkan data-data menjadi  $k$  rentang yang berbeda dari setiap *cluster* yang ada, yang mana  $k$  sendiri merupakan bilangan bulat positif.

Algoritma *K-means* identik dengan kesederhanaan dan keunggulannya dalam melakukan klasifikasi data dalam jumlah besar dan outlier dengan sangat cepat [4].

### F. Teknik *K-means*

*K-means* melakukan parameter input  $k$  dan kemudian melakukan pembagi satu set objek  $n$  menjadi  $k$ -*cluster*, selanjutnya akan menciptakan kesamaan intra-*cluster* yang tinggi tetapi kesamaan antar-*cluster* rendah. Kesamaan klaster diukur relatif terhadap rata-rata objek dalam klaster, yang dapat dilihat sebagai pusat gravitasi atau pusat massa klaster. Algoritma *K-means* kemudian mengulangi langkah-langkah berikut hingga stabilitas tercapai (hambatan tidak dapat dipindahkan):

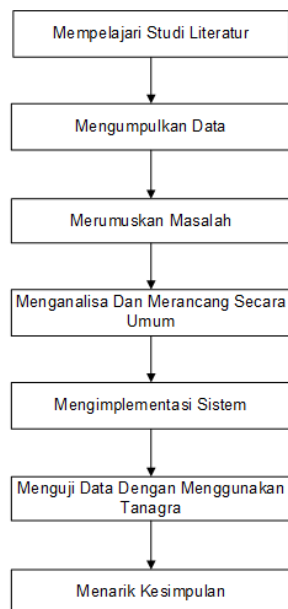
- Menentukan koordinat pusat setiap *cluster*,
- Tentukan jarak masing-masing benda terhadap koordinat pusatnya,
- Kelompokkan benda-benda tersebut menurut jarak minimumnya.

## III. METODE PENELITIAN

Metodologi penelitian ini meliputi langkah-langkah kerja yang harus dilakukan untuk memudahkan pengerjaan disertasi dan juga dapat dijadikan pedoman bagi peneliti dalam melakukan penelitian.

Penelitian adalah serangkaian kegiatan ilmiah yang ditujukan untuk memecahkan suatu masalah. Target dari penelitian yang dilakukan ialah menghasilkan penjelasan dan jawaban terhadap masalah yang diteliti serta menemukan solusi/ alternatif terbaik yang dapat digunakan untuk menyelesaikan sebuah persoalan/masalah.

Kerangka acuan diperlukan untuk langkah-langkah penelitian dilakukan secara terstruktur menciptakan langkah metodologi penelitian untuk memaksimalkan hasil yang diperoleh. Kerangka penelitian ini dapat dilihat pada gambar di bawah ini:



Gambar 1. Kerangka Kerja Penelitian

## 1. Studi Literatur

Pada tahap ini peneliti meninjau literatur tentang data mining, *clustering* dan algoritma *K-means*. Literatur yang dijadikan bahan rujukan antara lain: buku teks, jurnal, website, esai, *e-book* dan artikel akademik.

## 2. Pengumpulan Informasi

Pada tahap ini, data dikumpulkan dari CV melalui observasi langsung. Berkah Electronics, data yang dikumpulkan adalah data penjualan. Setelah data terkumpul dan diteliti, langkah selanjutnya adalah mengelompokkan data-data terkait hasil penjualan dan dari hasil pengelompokan data-data penjualan tersebut kemudian digunakan dalam proses menyelesaikan permasalahan yang akan dicari solusi terbaiknya.

## 3. Merumuskan Masalah

Setelah tahapan pengumpulan data telah selesai dilakukan, tahapan berikutnya ialah merumuskan pertanyaan-pertanyaan tentang permasalahan yang ada. Cara membuat keputusan tentang menentukan tingkat penjualan produk elektronik yang ada pada CV. Berkah Electronics dengan teknik algoritma *K-means* berdasarkan data penjualan yang ada.

## 4. Analisis dan Perencanaan secara umum

Pada tahapan ini dijelaskan metode analitik yang biasa digunakan dalam desain untuk menentukan produk elektronik terlaris.

## 5. Implementasi Sistem

Pada fase ini, program yang dibuat dengan aplikasi *Tanagra* dijalankan untuk menganalisis dan mengevaluasi hasil dari metode algoritma *K-means*. Pada tahap ini peneliti mengkaji literatur tentang data mining, *clustering* dan algoritma *K-means*. Literatur yang digunakan dari berbagai sumber diantaranya; buku teks, jurnal, website, buku, *e-book* dan buku akademik.

## 6. Menguji Data dengan *Tanagra*

Langkah selanjutnya setelah pemrograman adalah pengujian sistem. Pada fase ini, serangkaian pengujian dilakukan untuk memastikan bahwa sistem yang dirancang cocok dengan model yang dirancang. Langkah-langkah dalam mekanisme pengujian *data mining* adalah sebagai berikut:

### 1. Pemilihan informasi

Pemilihan (selection) data dilakukan dari koleksi data-data operasional harus dilakukan sebelum tahapan *data mining* pada *KDD*. Informasi yang dipilih dan digunakan dalam proses penambangan data disimpan kedalam file yang terpisah dari *database* operasional.

### 2. *Pretreatment*/pembersihan

Sebelum tahapan/proses *data mining* dapat dilakukan, dibutuhkan terlebih dahulu proses pembersihan terhadap data yang menjadi fokus *KDD*. Proses pembersihan tidak terbatas pada menghapus duplikat data, memeriksa data yang tidak konsisten, dan mengoreksi kesalahan pada data, seperti B. kesalahan ketik *KDD*, semacamnya. B. data atau informasi eksternal.

### 3. Transformasi/ Perubahan

Koding merupakan tahapan transformasi data yang dipilih sehingga data tersebut cocok untuk proses data mining. Proses pengkodean *KDD* merupakan proses kreatif dan sangat bergantung pada jenis/model data yang diambil dari database yang tersedia.

### 4. Interpretasi/Evaluasi

Model data output dari proses *data mining* harus bisa menghasilkan data dalam bentuk yang

mudah dimengerti oleh berbagai pihak yang memiliki kepedulian terhadap output data tersebut. Data yang dihasilkan tidak bisa terlepas dari setiap langkah/ tahapan yang menjadi bagian yang tak terpisahkan dari proses KDD yang dikenal dengan istilah rendering. Pada fase ini akan dipastikan apakah data output dihasilkan sesuai atau malah bertentangan dengan fakta/hipotesis yang sudah ada.

## 7. Membuat kesimpulan

Pada tahapan akhir penelitian akan ditarik kesimpulan yang tujuannya adalah untuk membandingkan hasil tahap implementasi sistem manual dengan sistem metode algoritma *K-means Clustering*.

## IV. HASIL DAN PEMBAHASAN

### A. Data Analysis

CV. Berkah Elektronik memiliki informasi tentang kegiatan penjualan perusahaan. Salah satunya adalah data-data transaksi yang terjadi di perusahaan dalam rentang waktu tertentu. Data-data tersebut terdiri dari beberapa atribut diantaranya: merk elektronik, jenis barang, nama barang, harga barang dan jumlah barang yang terjual.

Data digunakan sebagai atribut untuk melakukan pengolahan data. Pengertian *best seller* (Laris) adalah barang yang laku dalam jumlah banyak, sedangkan pengertian *no seller* adalah barang yang tidak laku (tidak Laris) atau memiliki volume penjualan di bawah rata-rata.

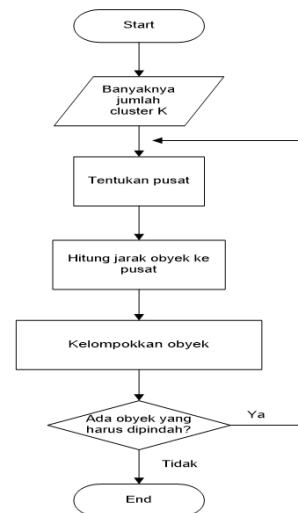
### B. Algoritma *K-means* untuk *Clustering Analysis*

*K-means* termasuk dalam metode pengelompokan partisi dari data mining, semua data harus dimasukkan dalam *cluster* tertentu dan semua data milik *cluster* tertentu dapat ditransfer ke *cluster* lain dalam satu langkah proses, di langkah berikutnya.

*K-means* membagi data menjadi  $K$  rentang terpisah, di mana  $K$  adalah bilangan bulat positif. Algoritma *K-means* dikenal dengan kesederhanaan dan kemampuannya untuk mengklasifikasikan data besar dan outlier dengan sangat cepat.

Berikut adalah flowchart dari algoritma *K-means* dan menjelaskan langkah-langkah dari

algoritma *K-means* dengan asumsi parameter input adalah jumlah record  $n$  data dan jumlah pusat inialisasi  $K=2$  sesuai dengan penelitian



Gambar 2. Flowchart Proses *K-means*

#### 1. Penentuan jumlah *cluster*

Pada tahap pertama ini, jumlah *cluster* ditentukan dengan menggunakan data yang diperoleh.

#### 2. Mendefinisikan pusat klaster pertama

Dalam menentukan  $n$  pusat klaster awal, dihasilkan bilangan acak yang mewakili urutan data masukan. Pusat awal *cluster* diperoleh dari data itu sendiri, tidak mendefinisikan titik baru, yaitu dengan mengacak pusat awal data.

#### 3. Hitung jarak objek ke pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster*, digunakan Euclidean distance, yang merupakan algoritma untuk menghitung jarak antara data dan pusat *cluster*: Ambil nilai data dan nilai pusat klaster, lalu hitung jarak Euclidean dengan masing-masing pusat klaster.

#### 4. Pengelompokan data objek

Jarak yang dihitung dibandingkan dan dipilih jarak terdekat antara data dengan pusat klaster, jarak ini menunjukkan bahwa data berada dalam kelompok yang sama dengan pusat klaster terdekat.

Algoritma pengelompokan data adalah sebagai berikut:

- a) Ambil nilai jarak dari pusat setiap *cluster* ke data.

- b) Cari nilai jarak terkecil.
- c) Mengelompokkan data berdasarkan pusat *cluster* dengan celah terdekat.

#### 5. Pembentukan pusat kluster baru

Dalam mendapatkan hasil pusat kluster baru dapat dilakukan dengan menghitung rata-rata anggota kluster dan pusat kluster. Pusat kluster baru tersebut akan dipergunakan dalam melakukan iterasi berikutnya ketika hasil yang diperoleh belum konvergen, menghentikan iterasi ketika jumlah maksimum iterasi yang dimasukkan oleh pengguna tercapai atau hasil yang diperoleh telah konvergen (pusat kluster baru sama dengan pusat kluster lama). Algoritma yang digunakan untuk menentukan pusat *cluster* adalah: Temukan jumlah anggota di setiap *cluster* dan kemudian hitung pusat baru menggunakan Formula 2.1

#### 6. Hitung jarak ke pusat *cluster*

Hitung jarak Euclidean dari semua data ke pusat baru (C1 dan C2) seperti pada langkah 2. Setelah kita mendapatkan hasil perhitungan, bandingkan hasil tersebut.

### C. Interpretasi / Evaluasi

Menafsirkan pola yang dihasilkan dari penambangan data dan kemudian memeriksa apakah pola atau data yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### D. Proses *clustering* menggunakan algoritma *K-means*

Pada fase ini, proses utama dilakukan dengan data penjualan merchandise berbasis merek CV. Berkah Elektronik, yaitu. segmentasi atau pengelompokan data penjualan barang terlihat pada laporan penjualan yang merupakan metode pengelompokan algoritmik *K-means*. Kumpulan data penjualan yang diperoleh dari laporan penjualan digunakan sebagai aplikasi algoritma *K-means* dalam penjualan barang. Eksperimen dilakukan dengan parameter berikut:

- a. Total *cluster*: 2
- b. Total data: 37
- c. Total atribut: 5

#### 1. Penentuan Jumlah *Cluster*

Menentukan jumlah *cluster* yang diperoleh dari data penjualan yang telah ada. Adapun jumlah *cluster* pada kasus ini adalah berjumlah tiga *cluster* yaitu merk, jenis, dan type.

#### 2. Penentuan Pusat *Cluster* Awal

Pusat awal *cluster* atau *centroid* didapatkan secara random atau acak berdasarkan data penjualan yang ada, untuk penentuan awal *cluster* diasumsikan:

- a) Pusat *cluster* 1: (25 , 6)
- b) Pusat *cluster* 2: (40 , 2)

#### 3. Perhitungan Jarak Object ke Pusat *Cluster*

Untuk mengukur jarak antara data dengan pusat *cluster* dapat digunakan Euclidean Distance

Dari 37 data disiapkan untuk menjadi sampel telah dilakukan pemilihan pusat awal *cluster* yaitu  $C1=(25 , 6)$ ,  $C2=(40 , 2)$ . Kemudian akan di hitung jarak dari sisa sampel data dengan pusat *cluster* yang dimisalkan dengan M.

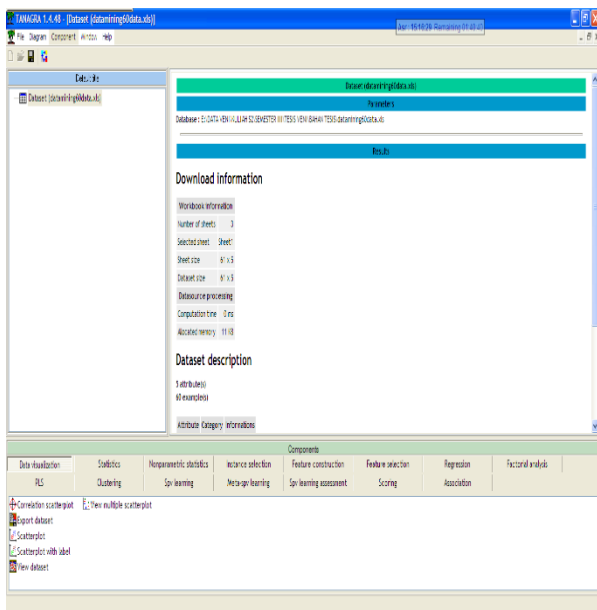
### E. System Implementation

Pengujian hasil analisis sangat penting untuk mengetahui dan memverifikasi apakah hasil analisis tersebut benar atau tidak. Menguji keakuratan hasil pengolahan data manual. Adapun tahapan pengujian yang dilakukan dalam menentukan tingkat penjualan barang elektronik pada CV. Berkah Elektronik pada pengujian sampel data penjualan dengan menggunakan salah satu software Tanagra 1.4.48. Dengan langkah-langkah sebagai berikut:

MERK	JENIS	TYPE	HARGA	JUMLAH
LG	TV	21FU58L	14	52
LG	TV	14FL7AB	8	13
LG	TV	14R020R	25	6
LG	AC	C-07LCS-2	40	2
LG	AC	S-03KCE-1.5	35	37
LG	DVD	DV-432	4.5	14
LG	KULKAS	CN-813T2C-1E	40	1
LG	KULKAS	CN-M93T2C-2	55	9
LG	KULKAS	CN-V12SL-1K	23	5
LG	MESIN CUCI	WV-613CB-1T	30	11
LG	MESIN CUCI	WV-6000-2T	20	3
PANASONIC	AC	C5-C10KCP-2	49.5	3
PANASONIC	BLENDER	MX-31G	3.95	1
PANASONIC	KIPAS ANGIN	FEL-409	8	8
PANASONIC	KULKAS	NB-A178X-1P	21	10
PANASONIC	KULKAS	NB-B203G-2P	37	2
PANASONIC	MESIN CUCI	NAW-70BC1-2T	18.5	19
PANASONIC	TV	TH-L32C10X	48	13
PANASONIC	TV	TH-L373D1D	77	2
POLYTRON	AC	JBL-688-1	29	9
POLYTRON	DISPENSER	PWC-107	15	10
POLYTRON	DVD	DVD-218B3	4.5	18
POLYTRON	KULKAS	PR-21188-2P	24.5	1
POLYTRON	KULKAS	PR-149VNB-1P	17	1
POLYTRON	MESIN CUCI	PAN-6205-2T	19	1
POLYTRON	TV	MX-1403	10	1
POLYTRON	TV	PS-213M	16	13
POLYTRON	TV	PS-296M	27	3
SHARP	AC	AR-ARCCY-1	27	5
SHARP	AC	AR-AP38BY-1.5	38	15
SHARP	AC	AR-AP38Z-2	40	2
SHARP	KULKAS	SJ-G170T2B-1	16.25	24
SHARP	KULKAS	SJ-G200UZS-2	29	11
SHARP	TV	21DN388	17	20
SHARP	TV	W2N300	24	5
TOSHIBA	MESIN CUCI	VM-B200B-1T	21	1
TOSHIBA	KULKAS	GR-R240RD-2P	30	1
PHILIPS	BLENDER	HR-2073	4	2
PHILIPS	SETRIKA	BB-114	2	2
RINAI	KOMPOR GAS	R9-522A	3	1
SAMSUNG	TV	CS-190M1ML	21	1
SAMSUNG	AC	AS-0582N-1	28	4
SONY	CAMERA DIGITAL	DSC-W180-B	18	9
SONY	CAMERA DIGITAL	DSC-W10-C	24	28
SONY	HANDYCAM	DCC-SX608	50	2
SONY	MINI COMPO	MBC-C21300D	61	1
SONY	PLAYSTATION	PS2-90006-1	14	8
TECHTRON	DVD	P-DVD9500	18	1
NIKON	CAMERA DIGITAL	S-33	23	1
OLYMPUS	CAMERA DIGITAL	FE-468	15	9
BBM	SPEAKER	SPK-550	23	1
DIGITEC	DVD	DR-1836	4	2
FUJIFILM	CAMERA DIGITAL	AC-C	18	1
FUJIFILM	CAMERA DIGITAL	SEM0-B	17	1
JVC	HANDYCAM	FR-D820	25.5	1
JVC	HANDYCAM	GC-M630	30	4
SHARP	VACUUM CLEANER	EC-8304R	8	4
MAYAKO	BLENDER	TMS-126	3	2
MAYAKO	DISPENSER	WV-188	2	1
MAYAKO	JUICER	JE-600	5	2

Gambar 3. Tampilan Database

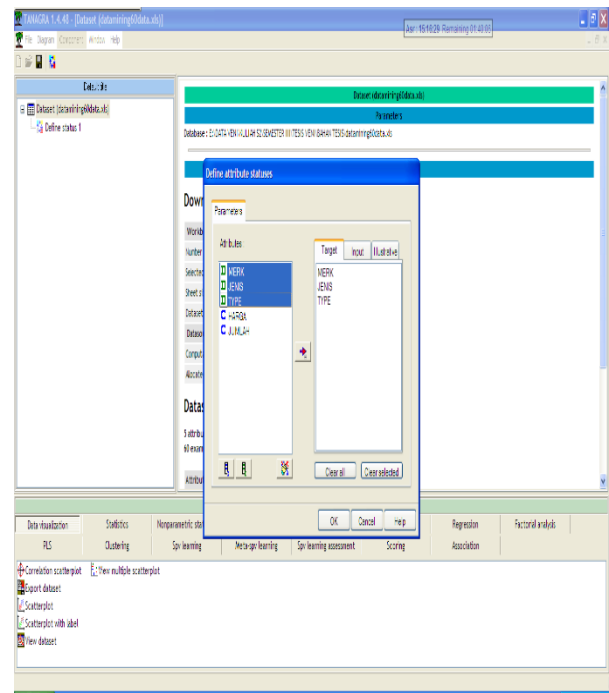
Setelah menentukan database yang akan digunakan selanjutnya kita mulai melakukan implementasi pada aplikasi Tanagra 1.4.48.



Gambar 4. Tampilan Dataset

Setelah ditampilkan dataset lalu kemudian tambahkan DEFINE STATUS, dan masukkan

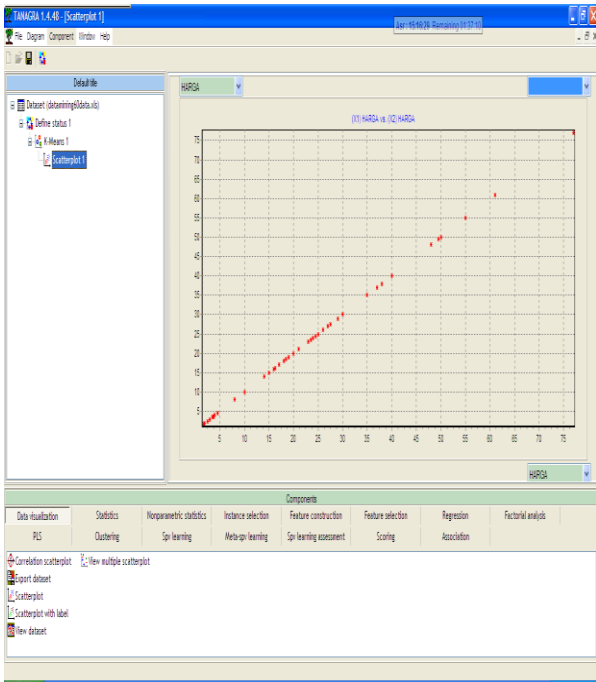
input HARGA dan JUMLAH kedalam Define Attribute Statuses, Dengan cara yang sama kemudian masukkan target MERK, JENIS dan TYPE kedalam Define Attribute Statuses



Gambar 5. Tampilan Define Status 1 Target

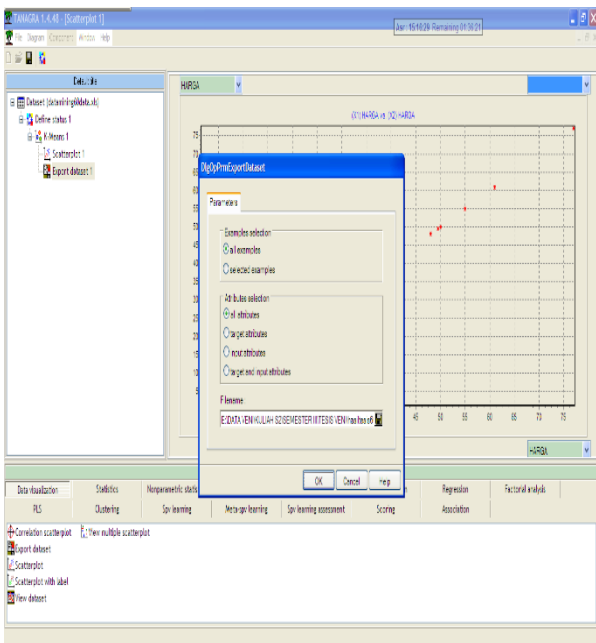
Pada tahapan berikutnya kita akan melihat *Clustering* pada menu yang ada di bagian bawah lalu pilih *K-means* kemudian tarik ke arah Define Status 1, kemudian *K-means* yang ada di bawah Define Status parameter, kemudian tampilkan layar *K-means* Parameters untuk menampilkan hasil *K-means* yang terdapat pada menu Data Visualization - Scatterplot;





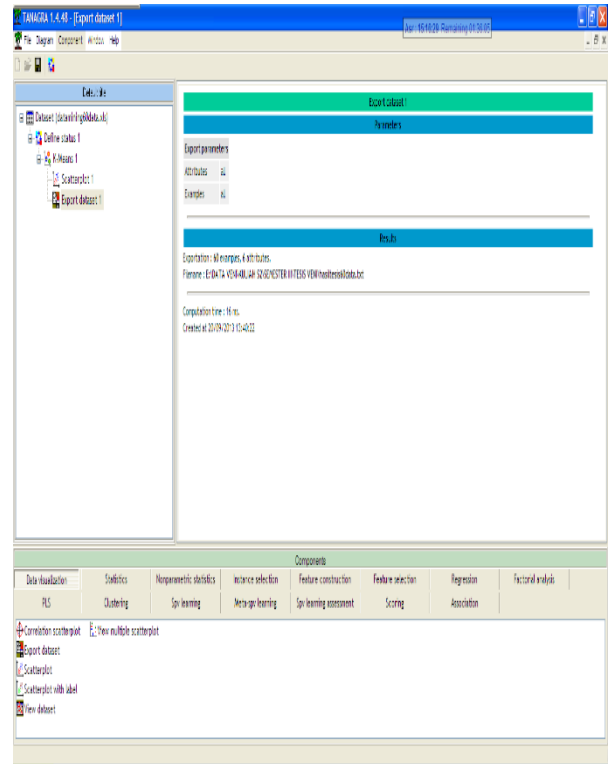
Gambar 6. Tampilan View Skatterflot

Langkah selanjutnya adalah menu Data Visualization pada menu akan di *Export Dataset 1* dan letakkan di bawah *K-means 1*, untuk melihat hasil *Export Dataset 1* didapat dengan *Eksport Dataset 1* lalu pilih parameter dan kemudian atur tempat penyimpanan hasil atau output dari proses yang telah dilakukan dimulai dari awal sampai akhir dengan hasil seperti gambar berikut:



Gambar 7. Tampilan DiGOpPmEksportDataset

Setelah ditentukan tempat penyimpanan dan jika ingin melihat kembali dimana hasil output nya tersimpan maka dapat dilakukan dengan mudah, dengan memview *Eksport Dataset 1* seperti gambar berikut:



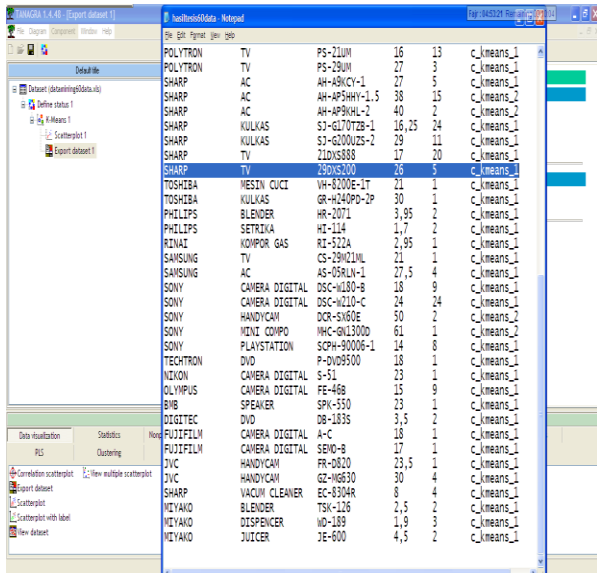
Gambar 8. Tampilan View Eksport Dataset 1

Tahapan berikutnya kita kan melihat hasil output dari data yang diproses:

MERK	JENIS	TYPE	HARGA	JUWILAH	Cluster_KMeans
LG	TV	21P08L	14	52	c_kmeans_1
LG	TV	14P07AB	8	13	c_kmeans_1
LG	TV	19LW20R	25	6	c_kmeans_1
LG	AC	C-07LCS-2	40	2	c_kmeans_2
LG	AC	S-05ECE-1.5	35	37	c_kmeans_2
LG	DVD	DIV-452	4,5	14	c_kmeans_1
LG	KULKAS	GM-B352YLC-1B	40	1	c_kmeans_2
LG	KULKAS	GM-M992YPC-2	55	9	c_kmeans_2
LG	KULKAS	GM-V212SL-1K	23	5	c_kmeans_1
LG	MESIN CUCI	WF-S15CR-1T	30	11	c_kmeans_1
LG	MESIN CUCI	WP-600N-2T	20	3	c_kmeans_1
PANASONIC	AC	CS-CL00KP-2	49,5	3	c_kmeans_2
PANASONIC	BLENDER	VN-31G	3,95	1	c_kmeans_1
PANASONIC	KIPAS ANGIN	FEU-409	3,7	8	c_kmeans_1
PANASONIC	KULKAS	NR-A17CX-1P	25	10	c_kmeans_1
PANASONIC	KULKAS	NR-B203G-2P	37	2	c_kmeans_2
PANASONIC	MESIN CUCI	NAW-70BC1-2T	18,5	19	c_kmeans_1
PANASONIC	TV	TH-L32C10X	48	13	c_kmeans_2
PANASONIC	TV	TH-L37X10	77	2	c_kmeans_2
POLYTRON	AC	PLC-686-1	29	9	c_kmeans_1
POLYTRON	DISPENSER	PUC-107	15	10	c_kmeans_1
POLYTRON	DVD	DVD-2183	4,5	18	c_kmeans_1
POLYTRON	KULKAS	PR-211SH-2P	24,5	1	c_kmeans_1
POLYTRON	KULKAS	PR-169YB-1P	17	1	c_kmeans_1
POLYTRON	MESIN CUCI	PAW-620S-2T	19	1	c_kmeans_1
POLYTRON	TV	VN-1403	10	1	c_kmeans_1
POLYTRON	TV	PS-21UM	16	13	c_kmeans_1
POLYTRON	TV	PS-29UM	27	3	c_kmeans_1
SHARP	AC	AH-A9NCV-1	27	5	c_kmeans_1
SHARP	AC	AH-A9HW-1.5	38	15	c_kmeans_2
SHARP	AC	AH-A9HW-2	40	2	c_kmeans_1
SHARP	KULKAS	S3-G170TB-1	16,25	24	c_kmeans_1
SHARP	KULKAS	S3-G200U-2S-2	29	11	c_kmeans_1
SHARP	TV	21DX588	17	20	c_kmeans_1



Gambar 9. Tampilan Output untuk 60 Record



Gambar 10. Tampilan Output untuk 60 Record (Lanjutan)

Di mana output yang dihasilkan pada gambar 5.14 adalah :

- 1) Jika *Cluster\_KMeans\_1* atau *c\_kmeans\_1* = maka tingkat penjualan = Laris
- 2) Jika *Cluster\_KMeans\_2* atau *c\_kmeans\_2* = maka tingkat penjualan = Tidak Laris.

### F. Pembahasan Hasil Pengujian

Dari tahapan pengujian yang telah dilaksanakan dengan mengikuti semua proses secara teknik manual dan dengan software Tanagra 1.4.48, dapat disimpulkan bahwa hasil pengujian sama dengan skor *centroid* dan hasil *clustering* dihasilkan oleh Tanagra 1.4. 48 software sama dengan menggunakan perhitungan manual. Begitu juga pengujian menggunakan seluruh data (60 record). hasil dari perhitungan manual dari 60 record data yang menghasilkan tiga iterasi dari perhitungan manual *Euclidean Distance* satu dan *Euclidean Distance* dua maka diperoleh jumlah *Cluster 1* sebanyak 12 record data, dan untuk *Cluster 2* sebanyak 28 record data.

Setelah perhitungan manual dilakukan pengujian dengan menggunakan software Tanagra 1.4.48 untuk membandingkan hasil yang diperoleh. Dalam melakukan pengujian pada software Tanagra 1.4.48 dari 60 record data penjualan maka diperoleh hasil untuk jumlah *Cluster 1* sebanyak 12 data, dan untuk *Cluster 2* sebanyak 28 jumlah data. Maka dapat dikatakan perhitungan manual dengan pengujian menggunakan seluruh data (60 record).

pada software Tanagra 1.4.48 memiliki hasil yang sama.

Hasil pengujian *cluster* pertama dari 60 data yang dilakukan memberikan tidak kurang dari 36 objek tersaji dalam tabel berikut:

Tabel I. Hasil *Cluster* Pertama

NO	MERK	JENIS	TYPE	HARGA	JML
1	LG	TV	21FU6RLR	14	52
2	LG	TV	14FU7ABF	8	13
3	LG	TV	19LH20RC	25	6
4	LG	DVD	DV-452	4,5	14
5	LG	KULKAS	GN-VL12SL-1K	23	5
6	LG	ME SIN CUCI	WF-15CR-1T	30	11
7	LG	ME SIN CUCI	WP-60IN-2T	20	3
8	PANASONIC	BLENDER	MX-J1G	3,95	1
9	PANASONIC	KIPAS ANGIN	FEU-409	3,7	8
10	PANASONIC	KULKAS	NR-A17KX-1P	23	10
11	PANASONIC	ME SIN CUCI	NAW-70BC1-2T	18,5	19
12	POLITRON	AC	PLE-686F-1	29	9
13	POLITRON	DISPENCER	PWC-107F	15	10
14	POLITRON	DVD	DVD-2183	4,5	18
15	POLITRON	KULKAS	PR-211SM-2P	24,5	1
16	POLITRON	KULKAS	PR-169VSB-1P	17	1
17	POLITRON	ME SIN CUCI	PAW-6T205-2T	19	1
18	POLITRON	TV	MIX-1403	10	1
19	POLITRON	TV	PS-21UM	16	13
20	POLITRON	TV	APS-29UM	27	3
21	SHARP	AC	AH-A9KCY-1	27	5
22	SHARP	KULKAS	SI-G170TZB-1	16,25	24
23	SHARP	KULKAS	SI-G200UZS-2	29	11
24	SHARP	TV	21DXS888	17	20
25	SHARP	TV	29DXS200	26	5
26	TOSHIBA	ME SIN CUCI	VH-8200E-1T	21	1
27	TOSHIBA	KULKAS	GR-H240PD-2P	30	1
28	PHILIPS	BLENDER	HR-2071	4	2
29	PHILIPS	SETRIKA	HI-114	2	2
30	RINAI	KOMPOR GAS	RI-522A	3	1
31	SAMSUNG	TV	CS-29M2IML	21	1
32	SAMSUNG	AC	AS-03RLN-1	28	4
33	SONY	CAMERA DIGITAL	DSC-W180-B	18	9
34	SONY	CAMERA DIGITAL	DSC-W210-C	24	24
35	SONY	PLAYSTATION	SCPH-90006-1	14	8
36	TECHTRON	DVD	P-DVD9500	18	1

Hasil *cluster* kedua saat dilakukan pengujian terhadap 60 data, menghasilkan 12 item terseaji pada tabel berikut;

Tabel II. Hasil *Cluster* Kedua

NO	MERK	JENIS	TYPE	HARGA	JML
1	LG	AC	CT-07LCS-2	40	2
2	LG	AC	ST-05ICE-1.5	35	37
3	LG	KULKAS	GN-B352YLC-IB	40	1
4	LG	KULKAS	GN-M392YPC-2	55	9
5	PANASONIC	AC	CS-C10KCP-2	49,5	3
6	PANASONIC	KULKAS	NR-B203G-2P	37	2
7	PANASONIC	TV	TH-L32C10X	48	13
8	PANASONIC	TV	TH-L37X10	77	2
9	SHARP	AC	AH-AP5HHY-1.5	38	15
10	SHARP	AC	AH-AP9KHL-2	40	2
11	SONY	HANDYCAM	DCR-SX60E	50	2
12	SONY	MINI COMPO	MHC-GN1300D	61	1

Algoritma *K-means* dianggap sebagai algoritma yang membantu dalam kelompok data yang propertinya dapat dikenali dengan jelas dan memudahkan pengguna untuk mengekstrak informasi dari data yang relevan.

## V. KESIMPULAN

Dari uraian dan penjelasan yang sudah dibahas pada penelitian bisa ditarik beberapa kesimpulan antara lain: Dengan algoritma *K-means*, penentuan tingkat penjualan produk elektronik lebih mudah dibandingkan dengan cara manual. Pengelompokan yang dihasilkan digunakan untuk menyimpulkan data penjualan barang elektronik. Berkah Electronics menggunakan algoritma *K-means* untuk menentukan produk mana yang berkinerja baik dan mana yang tidak. Pemilihan variabel dan atribut yang akan digunakan memiliki dampak yang signifikan terhadap data yang digunakan. Pengembangan sistem ini dari sistem manual atau data menjadi sistem komputerisasi dilakukan dari perspektif masalah yang timbul dari sistem yang lama. Berikut adalah saran yang dapat dipertimbangkan untuk masa depan; Pada penelitian ini, bereksperimen dengan salah satu metode *clustering data mining* yang ada yaitu algoritma *K-means*. Untuk kedepannya perlu

dikembangkan penelitian lebih mendalam untuk membandingkan hasil dan mencari tahu teori mana yang memberikan hasil yang baik. Perangkat lunak yang digunakan adalah *Tanagra 1.4.48*. Meskipun perangkat lunak ini mudah digunakan, Anda perlu memahami dan mengerti terlebih dahulu cara menggunakannya.

## REFERENSI

- [1] Kusriani & Emha Taufiq Luthfi (2019), "*Algoritma Data Mining*", Andi Yogyakarta, Yogyakarta
- [2] Santosa., Budi (2017), "*Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*", Graha Ilmu, Yogyakarta, Yogyakarta
- [3] Hermawati, Fajar Astuti (2013), "*Data Mining*", Andi Yogyakarta, Yogyakarta
- [4] Wijaya, Arim (2019), "*Jurnal Analisa Algoritma Means Untuk Sistem Pendukung Keputusan Penjurusan Siswa Di MAN BINONG SUBANG*"
- [5] Dana Sulistiyo Kusumo, Moch. Arief Bijaksana, Dhinta Darmantoro "*Jurnal Data mining Dengan Algoritma Apriori Pada RBMS Oracle*"
- [6] Irdiansyah, Enur (2013), "*Jurnal Penerapan Data mining Pada Penjualan Produk Minuman di PT.Pepsi Cola Indobeverages Menggunakan Metode Clustering*"
- [7] Mewati Ayub (2017), "*Jurnal Proses Data mining Dalam Sistem Pembelajaran Berbantuan Komputer*"
- [8] Wakidah, Nur (2015), "*Jurnal Clustering Menggunakan K-means Algorithm (K-means Alorithm Clustering)*"
- [9] Pramudiono, Iko (2013), "*Pengantar Data mining : Menambang Permata Pengetahuan di Guning Data*"
- [10] Andayani, Sri (2017), "*Pembentukan Cluster ke Dalam Knowledge Discovery in Database Dengan Algoritma K-means*"

- [11] Sunjana (2014), “*Jurnal Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree*”
- [12] Agusta, Yudi (2015), “*Jurnal K-means Penerapan, permasalahan dan Metode Terkait*”
- [13] M. Abinaya and R. U. Devi, “Intelligent Vehicle Control Using Wireless Embedded System in Transportation System Based On GSM and GPS Technology,” vol. 3, no. 9, pp. 244–258, 2014.
- [14] D. Hartanti, R. N. Aziza, and P. C. Siswipraptini, “Optimization of smart traffic lights to prevent traffic congestion using fuzzy logic,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 17, no. 1, pp. 320–327, 2019, doi: 10.12928/TELKOMNIKA.v17i1.10129.
- [15] P. C. Siswipraptini, W. H. Martono, and D. Hartanti, “Reducing a congestion with introduce the greedy algorithm on traffic light control,” *J. Phys. Conf. Ser.*, vol. 974, no. 1, 2018, doi: 10.1088/1742-6596/974/1/012013.
- [16] J. Y. Kim, H.-J. Lee, J.-Y. Son, and J.-H. Park, “Smart Home Web of Objects-based IoT Management Model and Methods for Home data mining,” 2015.
- [17] S. Ghosh, “Smart homes: Architectural and engineering design imperatives for smart city building codes,” in *International Conference on Technologies for Smart City Energy Security and Power: Smart Solutions for Smart Cities, ICSESP 2018 - Proceedings*, 2018, vol. 2018-Janua, pp. 1–4, doi: 10.1109/ICSESP.2018.8376676.
- [18] D. E. Paul, “Smart Energy Meter Using Android Application And Gsm Network,” *Int. J. Eng. Comput. Sci.*, vol. 05, no. 16058, pp. 16058–16063, 2016, doi: 10.18535/ijecs/v5i3.36.
- [19] B. Lashkari, Y. Chen, and P. Musilek, “Energy management for smart homes-state of the art,” *Applied Sciences (Switzerland)*, vol. 9, no. 17. 2019, doi: 10.3390/app9173459.
- [20] M. Samarth Pandit, M. Snehamandhre, and M. Meghananichal, “Smart Energy Meter using internet of Things (IoT),” *J. Eng. Res. www.vjer.in*, vol. 1, no. 2, pp. 222–229, 2017.
- [21] Y. S. Tak, J. Kim, and E. Hwang, “Hierarchical querying scheme of human motions for smart home environment,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 7, pp. 1301–1312, Oct. 2012, doi: 10.1016/j.engappai.2012.03.020.
- [22] S. Z. Reyhani and M. Mahdavi, “User authentication using neural network in smart home networks,” *Int. J. Smart Home*, vol. 1, no. 2, pp. 147–154, 2007.