

## **Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means**

Yahya Novi Andi Cuhwanto<sup>1</sup>; Dewi Agushinta R.<sup>2</sup>

<sup>1</sup>Program Pasca Sarjana Magister Manajemen Sistem Informasi, Universitas Gunadarma

<sup>2</sup>Sistem Informasi Fakultas Ilmu Komputer dan Teknologi, Universitas Gunadarma

<sup>1</sup> yahya\_novi@gmail.com

### **ABSTRACT**

*Data mining is an important process of mining information from data. The data mining revolution and the concept of Electronic Commerce (E-Commerce) are growing rapidly, many companies interest in increasing online transactions. From this data can be collected into information that will be useful for making a decision or action to be taken. The number of customers making transactions will show customer loyalty in E-Commerce, even from companies that can give gifts to potential customers. The company realizes the importance of the relationship between loyal customers and the company's business success. The stages carried out by researchers are data collection, initial data processing, data processing, and research results. The technique of collecting data is secondary data, already available on E-Commerce pasarayastore.com. The data used include Name, Valid orders, and Money spent (IDR) from January 2018 to June 2021. In this study, data mining analysis was carried out using K-Means algorithm clustering techniques. The K-Means algorithm is used to retrieve data and classify customers who make transactions the most frequently. The application used as a tool for testing transaction data is Tanagra. The result of this research is the grouping of data into categories of customers who frequently, moderately, and rarely transact.*

**Keywords:** Data mining, Clustering, Customer, K-Means

### **ABSTRAK**

*Data mining merupakan suatu proses penambangan informasi penting dari suatu data. Revolusi data mining dan konsep Electronic Commerce (E-Commerce) berkembang pesat, menumbuhkan minat banyak perusahaan untuk meningkatkan transaksi online. Dari data tersebut bisa dikumpulkan menjadi informasi-informasi yang akan berguna untuk membuat suatu keputusan atau tindakan yang akan diambil. Banyaknya pelanggan melakukan transaksi, maka akan terlihat kesetiaan pelanggan pada E-Commerce, bahkan dari perusahaan bisa memberikan hadiah kepada pelanggan yang potensial. Perusahaan menyadari akan pentingnya hubungan antara pelanggan yang loyal dengan keberhasilan bisnis perusahaan. Tahapan yang dilakukan adalah pengumpulan data, pengolahan data awal, pengolahan data dan hasil penelitian. Teknik pengumpulan data dilakukan secara sekunder, mengambil data yang telah tersedia pada E-Commerce pasarayastore.com. Data yang digunakan antara lain Name, Valid orders dan Money spent (IDR) dengan periode Januari 2018 sampai Juni 2021. Pada penelitian ini analisis data mining dilakukan dengan teknik clustering menggunakan algoritma K-Means. Algoritma K-Means akan digunakan untuk mengambil data dan mengelompokkan pelanggan yang paling sering melakukan transaksi. Aplikasi yang akan digunakan sebagai alat bantu pengujian data transaksi adalah Tanagra. Hasil dari penelitian ini adalah mengelompokkan data dalam kategori pelanggan yang sering, sedang dan jarang bertransaksi.*

**Kata kunci:** Clustering, Data mining, K-Means, Pelanggan

## 1. PENDAHULUAN

Bagian Pendahuluan membahas latar belakang masalah, tinjauan pustaka secara ringkas, maksud dan tujuan riset dilakukan. Pendahuluan sebaiknya berisi perkembangan penelitian terdahulu (*state of the art*) untuk membandingkan dengan penelitian yang dilakukan saat ini sehingga tampil kesenjangan antara teori atau hasil penelitian terdahulu dengan keadaan saat ini atau yang diharapkan. Pada Data Mining terdapat proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [1]. Algoritma K-Means adalah suatu metode penganalisis data atau metode data mining yang melakukan proses pemodelan tanpa supervisi dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi [2]. Metode K-Means merupakan salah satu metode dalam fungsi *clustering* atau pengelompokan. *Clustering* mengacu pada pengelompokan data, observasi atau kasus berdasarkan kemiripan objek yang diteliti. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain [3]. *Clustering* banyak digunakan di berbagai bidang seperti kesehatan, ekonomi, dan teknologi. Pengelompokan status kesehatan gizi balita pada Desa Kembang Songo menggunakan K-Means [4]. Pengujian *clustering* dengan algoritma K-Means meningkatkan penjualan terhadap data transaksi penjualan paket Telkomsel, diperoleh persentase kesesuaian 100% dibanding dengan *clustering* manual [5]. Media internet telah dijadikan sarana perantara dengan meningkatkan jumlah penjualan setiap tahunnya [6]. Penerapan algoritma K-Means *clustering* produk online shop untuk mengelompokkan produk menjadi kategori jumlah stok banyak, sedang dan sedikit berdasarkan transaksi penjualan [7].

Gaya hidup masyarakat saat ini ikut berubah karena pengaruh dari perkembangan teknologi tersebut, salah satu yang paling mencolok dari perkembangan teknologi tersebut adalah gadget dan kecenderungan beraktivitas di dunia maya seperti berbelanja secara online atau lebih sering disebut dengan belanja online [8]. Pada E-Commerce harga lebih transparan dan pemasaran dapat dilakukan secara global [9]. Berdasarkan latar belakang ini, Penulis akan melakukan proses analisis penentuan pelanggan potensial melalui pola-pola dari kumpulan data besar yang dimiliki pasarayastore.com dengan menggunakan algoritma K-Means.

Tujuan dari penelitian ini adalah melakukan analisis terhadap data transaksi pelanggan pada E-Commerce sehingga mendapatkan informasi pelanggan potensial dengan menggunakan algoritma K-Means. Menerapkan algoritma pengelompokan K-Means tradisional dan jarak pengukuran Euclidean dari kesamaan dipilih untuk digunakan dalam analisis nilai [10].

Laporan pelanggan juga membantu E-Commerce pasarayastore.com dapat melakukan pemilihan pelanggan potensial yang tepat sasaran untuk diberikan hadiah. Hasil penelitian ini diharapkan dapat memberikan pelayanan dan penjualan dengan cara yang mudah, respon cepat serta adanya dukungan sumber daya yang baik sehingga berimbas pada peningkatan pelayanan. Pemasaran merupakan suatu fungsi organisasi dan seperangkat proses untuk menciptakan, mengkomunikasikan, dan menyerahkan nilai kepada pelanggan serta mengelola hubungan pelanggan dengan cara yang menguntungkan organisasi dan para pemilik sahamnya [11].

## 2. METODE

### 2.1. Pengumpulan Data

Teknik pengumpulan data dilakukan secara sekunder, mengambil data secara langsung dari pasarayastore.com. Data yang didapat adalah *name*, *Valid orders*, dan *Money spent* (IDR).

Pengumpulan data transaksi penjualan pasarayastore.com pada periode Januari 2018 sampai Juni 2021, ditunjukkan di tabel 1.

**Tabel 1.** Data Transaksi Awal Pelanggan

<i>Name</i>	<i>Valid orders</i>	<i>Money spent (IDR)</i>
Paladima Rosi	0	0
Christianty Jesica	0	0
Ayu Lusitania	0	0
---	---	---
Tirta Elsa	4	6768957
Soewondo Seruni	3	10230069
Febrina Febrina	37	10582787

## 2.2. Data Preprocessing

Data preprocessing yaitu dilakukan pembersihan dan persiapan data, untuk menghilangkan kosistensi data, data tidak lengkap dan redundant data yang terdapat pada data awal. Normalisasi digunakan untuk menghilangkan data yang berlebihan dan memastikan bahwa *cluster* kualitas terbaik dihasilkan yang dapat meningkatkan efisiensi algoritma *clustering* [12]. Data preprocessing dengan melakukan penghapusan data transaksi pelanggan yang tidak pernah melakukan *Valid orders* atau bernilai 0.

## 2.3. Pengumpulan Data

Tahap ini akan dilakukan perubahan data yang memiliki tipe data yang tidak bisa diolah secara matematis menjadi data yang bisa diolah. Proses K-Means dilakukan transformasi data alpha numeric (teks) menjadi numerik. Pada kolom *Name* yang datanya berupa teks, diubah menjadi numerik dengan dijadikan nomor urut, seperti tabel 2.

**Tabel 2.** Transformasi Data

<i>Name</i>	<i>Valid orders</i>	<i>Money spent (IDR)</i>
1	1	4500
2	1	16000
3	3	18230
---	---	---
1481	4	6768957
1482	3	10230069
1483	37	10582787

## 2.4. Algoritma K-Means

*Clustering* adalah salah satu metode dalam data mining dengan tujuan utama mencari pola dari data untuk kemudian mengelompokkan data tersebut berdasarkan pola yang didapatkan, data dengan pola yang sama akan masuk ke *cluster* yang sama [13]. Algoritma K-Means adalah algoritma pada *clustering* yang melakukan pengelompokan dengan membandingkan jarak tiap data ke *centroid* tiap *cluster* dan nilai *centroid* didapatkan dari rerata tiap *cluster* [14].

Dalam penelitian mengungkapkan langkah-langkah pengerjaan algoritma K-Means, yaitu [15]:

1. Penentuan pusat *cluster* awal

Dalam penentuan nilai N dibuat pusat *cluster* awal, dilakukan pembangkitan bilangan random yang mempresentasikan urutan data input pusat awal *cluster*, bukan dengan menentukan titik baru, yaitu dengan mengacak pusat awal dari data.

2. Perhitungan jarak dengan pusat *cluster*  
Untuk mengukur jarak antara data dengan pusat *cluster* digunakan rumus jarak Euclidean. Algoritma dengan perhitungan jarak data dengan pusat *cluster*, yaitu:
  - a. Ambil nilai data dan nilai pusat *cluster*.
  - b. Hitung jarak Euclidean data dengan tiap pusat *cluster*.
3. Pengelompokkan data  
Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini akan menunjukkan data tersebut berada dalam suatu kelompok dengan pusat *cluster* terdekat. Algoritma pengelompokan data:
  - a. Ambil nilai jarak tiap pusat *cluster* dengan data.
  - b. Cari nilai jarak terkecil.
  - c. Kelompokkan data dengan pusat *cluster* yang memiliki jarak terkecil.
4. Penentuan pusat *cluster* baru
5. Untuk mendapatkan pusat *cluster* baru bisa dihitung dari rerata nilai anggota *cluster* dan pusat *cluster*. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi ini akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh user atau hasil dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).  
Algoritma penentuan pusat *cluster* baru yaitu:
  - a. Cari jumlah anggota tiap *cluster*.
  - b. Hitung pusat baru dengan rumus.

$$\text{Pusat } cluster \text{ baru} = \frac{x_1+x_2+x_3,\dots,x_n}{\text{jumlah}+1}$$

(1)

Dimana  $x_1, x_2, x_3, \dots, x_n$  adalah anggota *cluster*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Algoritma K-Means

Pada penggunaan algoritma K-Means akan melakukan pengulangan tahapan hingga terjadi kestabilan. Pengulangan dilakukan lebih dari satu kali iterasi dalam pengujian menggunakan algoritma K-Means pada data transaksi pelanggan dengan tahapan sebagai berikut:

1. Tahap awal dalam proses *clustering* K-Means adalah menentukan jumlah *cluster*. Penentuan *cluster* yang dibuat adalah 3 *cluster* yaitu *cluster* pertama (C1), *cluster* kedua (C2) dan *cluster* ketiga (C3).
2. Tahap kedua ini menentukan nilai pusat *cluster* (*centroid*) awal secara acak dari data transaksi pelanggan. Kriteria *centroid* diambil dari atribut-atribut data transaksi pelanggan, *centroid* kriteria 1 adalah *Valid orders* dan *centroid* kriteria 2 adalah *Money spent* (IDR). Diambil *record* ke 3 sebagai C1, *record* ke 700 sebagai C2 dan *record* ke 1477 sebagai C3 sebagai *centroid* awal dengan kriteria:  
C1 = (3:18230)

$$C2 = (1;269000)$$

$$C3 = (7;5514661)$$

Nilai pusat awal *cluster* ditentukan secara acak, maka hasil dari *centroid* awal berubah setiap kali melakukan percobaan proses *clustering*. Ada kemungkinan *centroid* awal yang dihasilkan pusat jauh berbeda dengan *cluster* yang terbaik, sehingga kemungkinan akan terjadi proses iterasi yang banyak untuk mencapai konvergen.

3. Menghitung jarak setiap data ke setiap *centroid* dengan menggunakan rumus jarak Euclidean. Berikut perhitungan menggunakan data pada tabel 3 dengan *centroid* awal yang sudah ditentukan:

Jarak data pertama ke pusat *cluster* pertama

$$d_{(1,1)} = \sqrt{(1 - 3)^2 + (4500 - 18230)^2} = 13730$$

Jarak data pertama ke pusat *cluster* kedua

$$d_{(1,2)} = \sqrt{(1 - 1)^2 + (4500 - 269000)^2} = 264500$$

Jarak data pertama ke pusat *cluster* ketiga

$$d_{(1,3)} = \sqrt{(1 - 7)^2 + (4500 - 5514661)^2} = 5510161$$

Jarak data kedua ke pusat *cluster* pertama

$$d_{(2,1)} = \sqrt{(1 - 3)^2 + (16000 - 18230)^2} = 2230$$

Jarak data kedua ke pusat *cluster* kedua

$$d_{(2,2)} = \sqrt{(1 - 1)^2 + (16000 - 269000)^2} = 253000$$

Jarak data kedua ke pusat *cluster* ketiga

$$d_{(2,3)} = \sqrt{(1 - 7)^2 + (16000 - 5514661)^2} = 5498661$$

Jarak data 1482 ke pusat *cluster* pertama

$$d_{(1482,1)} = \sqrt{(3 - 3)^2 + (10230069 - 18230)^2} = 10211839$$

Jarak data 1482 ke pusat *cluster* kedua

$$d_{(1482,2)} = \sqrt{(3 - 1)^2 + (10230069 - 269000)^2} = 9961069$$

Jarak data 1482 ke pusat *cluster* ketiga

$$d_{(1482,3)} = \sqrt{(3 - 7)^2 + (10230069 - 5514661)^2} = 4715408$$

Jarak data 1483 ke pusat *cluster* pertama

$$d_{(1483,1)} = \sqrt{(37 - 3)^2 + (10582787 - 18230)^2} = 10564557$$

Jarak data 1483 ke pusat *cluster* kedua

$$d_{(1483,2)} = \sqrt{(37 - 1)^2 + (10582787 - 269000)^2} = 10313787$$

Jarak data 1483 ke pusat *cluster* ketiga

$$d_{(1483,3)} = \sqrt{(37 - 7)^2 + (10582787 - 5514661)^2} = 5068126$$

**Tabel 3.** Jarak Antara *Centroid* Pada Iterasi 1

Name	Valid orders	Money spent (IDR)	C1	C2	C3
1	1	4500	13730	264500	5510161
2	1	16000	2230	253000	5498661
3	3	18230	0	250770	5496431
---	---	---			
1481	4	6768957	6750727	6499957	1254296
1482	3	10230069	10211839	9961069	4715408
1483	37	10582787	10564557	10313787	5068126

4. Mengelompokan data pada *cluster* terdekat. Setiap data dialokasikan ke suatu *cluster* berdasarkan jarak terdekat (hasil nilai terkecil) dari pusat *cluster*. Tabel 3 merupakan hasil perhitungan jarak setiap pelanggan dengan masing-masing *cluster*. Berdasarkan hasil tabel 3, pelanggan dengan nomor 1 memiliki jarak terdekat dengan *cluster* pertama dengan nilai *centroid* yaitu 13730, sehingga pelanggan nomor satu masuk ke dalam *cluster* pertama. Pada iterasi pertama diperoleh hasil 239 data masuk ke dalam *cluster* pertama, 1221 data masuk ke dalam *cluster* kedua dan 23 data masuk ke dalam *cluster* ketiga, seperti tabel 4.

**Tabel 4.** Pengelompokan *Centroid* Iterasi 1

Name	Valid orders	Money spent (IDR)	C1	C2	C3
1	1	4500	1	0	0
2	1	16000	1	0	0
3	3	18230	1	0	0
---	---	---			
1481	4	6768957	0	0	1
1482	3	10230069	0	0	1
1483	37	10582787	0	0	1
			239	1221	23

5. Menentukan nilai pusat *cluster* baru. Perhitungan *centroid* baru dengan menggunakan persamaan (1). Nilai *centroid* yang baru digunakan untuk melakukan iterasi selanjutnya. *Centroid* baru dihitung dari rerata nilai anggota *cluster* dan nilai pusat *cluster* masing-masing, maka perhitungan pada penelitian ini, sebagai berikut:

$$C1(\text{Valid orders}) = \frac{(1 + 1 + 3 + \dots + 1)}{239} = 1,0084$$

$$C1(\text{Money spent}) = \frac{(4500 + 16000 + 18230 + \dots + 143250)}{239} = 99074,7741$$

$$C2(\text{Valid orders}) = \frac{(1 + 1 + 1 + \dots + 1)}{1221} = 1,1523$$

$$C2(\text{Money spent}) = \frac{(144005 + 144090 + 144500 + \dots + 2724243)}{1221} = 418385,6$$

$$C3(\text{Valid orders}) = \frac{(4 + 3 + 12 + \dots + 37)}{23} = 9,0435$$

$$C3(\text{Money spent}) = \frac{(2909836 + 2957326 + \dots + 10582787)}{23} = 4819701,65$$

Hasil *centroid* baru pada iterasi pertama adalah:

$$C1 = (1,0084;99074,7741)$$

$$C2 = (1,1523;418385,644)$$

$$C3 = (9,0435;4819701,65)$$

6. Tahap 3 sampai 5 akan terus berulang apabila masih ada data yang berpindah *cluster* atau ada perubahan nilai pada *centroid*.
7. Proses iterasi ini akan berhenti jika hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama). Pada penelitian ini proses berhenti pada iterasi ke 11 dengan nilai *centroid*:

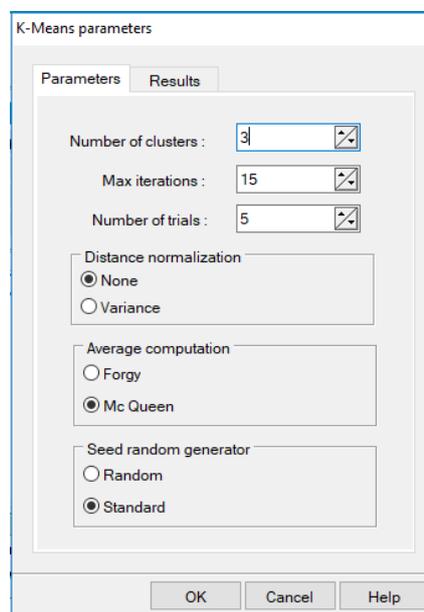
$$C1 = (11,0833;6258701,667)$$

$$C2 = (1,0969;319047,216)$$

$$C3 = (2,7059;1803723,7647)$$

### 3.2. Analisis dengan Tanagra

Pada analisis ini digunakan perangkat lunak Tanagra untuk penyelesaian dari algoritma K-Means. Langkah pertama yang harus dilakukan adalah mempersiapkan data transaksi pelanggan yang sudah ditransformasi dalam format .xls untuk diimpor ke dalam aplikasi Tanagra. Variabel parameter K-Means yang diberi nilai adalah *Number of clusters* (jumlah *cluster* yang sudah ditentukan) dengan nilai 3, dan *Max iterations* dengan nilai 15, setelah itu mengklik tombol Ok, seperti gambar 1.



**Gambar 1.** Parameter K-Means di Tanagra

Selanjutnya dipilih *Execute* pada pilihan menu K-Means 1, maka aplikasi Tanagra memproses algoritma K-Means sesuai dengan variabel parameter yang dimasukkan. Hasil dari *Execute* menu K-Means seperti gambar 2 dan gambar 3.

Clusters	3	
Cluster	Description	Size
cluster n°1	c_kmeans_1	12
cluster n°2	c_kmeans_2	1403
cluster n°3	c_kmeans_3	68

Gambar 2. Hasil Cluster di Tanagra

Cluster sudah ditentukan sebelumnya sebanyak 3. Gambar 2 menjelaskan bahwa pada cluster 1 memiliki data sebanyak 12, cluster 2 memiliki data sebanyak 1403 dan cluster 3 memiliki data sebanyak 68 dengan total semua data sebanyak 1483.

### Cluster centroids

Attribute	Cluster n°1	Cluster n°2	Cluster n°3
Valid orders	11.083333	1.096935	2.705882
Money spent (IDR)	6258701.666667	319047.215966	1803723.764706

Gambar 3. Hasil Centroid di Tanagra

Pada setiap cluster terdapat titik pusat (*centroid*) yang merepresentasikan cluster tersebut. Gambar 3 menjelaskan bahwa *centroid Valid orders* pada cluster 1 adalah 11, *centroid Valid orders* pada cluster 2 adalah 1 dan *centroid Valid orders* pada cluster 3 adalah 2,7. *Centroid Money spent (IDR)* pada cluster 1 adalah 6258702, *centroid Money spent (IDR)* pada cluster 2 adalah 319047 dan *centroid Money spent (IDR)* pada cluster 3 adalah 1803723.

## 4. KESIMPULAN DAN SARAN

Analisis terhadap data transaksi pelanggan E-Commerce pasarayastore.com untuk mendapatkan informasi pelanggan potensial dengan menggunakan algoritma K-Means berhasil dilakukan. Berdasarkan dari hasil perhitungan *clustering* K-Means pelanggan terkelompok menjadi 3 cluster, yaitu 12 pelanggan dengan rata-rata 11 kali bertransaksi (*cluster 1*), 1403 pelanggan dengan rata-rata 1 kali bertransaksi (*cluster 2*), dan 68 pelanggan dengan rata-rata 3 kali bertransaksi (*cluster 3*). Pelanggan potensial berhasil didapatkan, yaitu yang memiliki rata-rata transaksi dan uang yang dibelanjakan terbanyak pada cluster pertama (C1). Pelayanan dan penjualan menjadi mudah, respon cepat serta adanya dukungan sumber daya yang baik berimbas pada peningkatan pelayanan terhadap pelanggan potensial.

Pada perhitungan jarak Euclidean nilai kriteria *Valid orders* dan *Money spent (IDR)* dengan jarak antara titik kriteria sangat jauh yaitu bernilai satuan ke ribuan (1:4500), sehingga perhitungan lebih dominan pada nilai yang terbesar. Perhitungan jarak Euclidean dengan jarak antara titik kriteria dekat akan menghasilkan nilai yang tidak terlalu jauh antara kedua kriteria tersebut. Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang penetapan pelanggan potensial. Diharapkan penggunaan algoritma K-Means tidak hanya diterapkan pada pelanggan potensial saja, dapat juga pada jenis produk yang paling diminati pelanggan.

**DAFTAR PUSTAKA**

- [1] Gunadi G dan Sensuse I. (2012). Penerapan Metode Data Mining Market Base Analysis terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern-Growth (FP- Growth): Studi Kasus Percetakan PT. Gramedia. Volume 4 No. 1, Halaman 121.
- [2] Khomarudin, Agus Nur. (2016). Teknik Data Mining : Algoritma K-Means Clustering. Ilmukomputer.
- [3] Larose, D. T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- [4] Duhita Windha Mega Pradnya. (2015). Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita. *J. Informatika*, Vol. 15, No 2, Desember 2015.
- [5] Handoko, Suhandio. (2020). Implementasi Data Mining untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi dan Rekayasa* Volume 25 No. 1, April 2020.
- [6] Kim, J. and Forsythe, S., (2010). Factors affecting adoption of product virtualization technology for online consumer electronics shopping. *International Journal of Retail & Distribution Management*, 38 (3), 190–204.
- [7] Muningsih, Elly. (2015). Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang *Jurnal Bianglala Informatika* Vol 3 No 1, Maret 2015.
- [8] Kusuma L.A , dan Made. (2015). Pengaruh Persepsi Harga, Kepercayaan, dan Orientasi Belanja Terhadap Niat Beli Secara Online (studi pada produk fashion online di kota Denpasar). *E- Jurnal Manajemen Unund*, Vol.4, No.8. ISSN : 2302-8912.
- [9] Laudon, K. & Traver, C., 2016. *E-commerce: business, technology, society*. 12th ed. s.l.:Pearson.
- [10] S. Khatri dan K. Garg, “Document Clustering Using Improved K-Means Algorithm”, ISSN 2091-2730, *International Journal of Engineering Research and General Science* Volume 4, Issue 3, May-June, 2016.
- [11] Kotler dan Keller, 2007. *Manajemen Pemasaran*, Edisi 12, Jilid 1, PT. Indeks, Jakarta.
- [12] Deepali Virmani, Shweta Taneja, Geetika Malhotra, “Normalization based K means Clustering Algorithm” , ISSN: 2349-6495, *International Journal of Advanced Engineering Research and Science (IJAERS)*, [Vol-2, Issue-2, Feb.- 2015].
- [13] Adolfsson, A., Ackerman, M. & Brownstein, N., 2018. To Cluster, or Not To Cluster: An Analysis of Clusterability Methods. *Pattern Recognition*, Issue 88, pp. 13-26
- [14] Chayangkoon, N. & Srivihok, A., 2016. Two Step Clustering Model for K-Means Algorithm. s.l., *International Conference on Network, Communication and Computing*, pp. 213-217.
- [15] B. Santosa, *Data mining (Teknik pemanfaatan data untuk keperluan bisnis)*. Yogyakarta: Graha Ilmu, 2007.