

## **ANALISIS CLUSTER DENGAN MENGGUNAKAN K-MEANS UNTUK PENGELOMPOKKAN ONLINE CUSTOMER REVIEWS (OCR) PADA ONLINE MARKETPLACE**

**Rena Nainggolan<sup>1</sup>, Fenina A.T Tobing<sup>2</sup>**

<sup>1</sup>Universitas Methodist Indonesia

<sup>2</sup>Universitas Multimedia Nusantara

[renanain66olan@gmail.com](mailto:renanain66olan@gmail.com)

### **ABSTRACT**

Technological advances at this time are very influential on people's shopping culture, plus during the current pandemic, it has resulted in an increasing number of people shopping for daily necessities online. There are many conveniences offered in online shopping that make people switch to using these facilities. Besides the advantages of online shopping, there are also some disadvantages of online shopping, including the rise of online sales fraud such as goods not being shipped, damaged goods, items not as ordered, and much more. For this reason, in conducting online transactions, trust is needed between the seller and the buyer, and one of the factors that greatly affects the prospective buyer is to know the history of the seller, namely by looking at the reviews given by the buyer on the seller's homepage which is called Online Customers Reviews (OCR) . OCR is considered to be very influential on customer buying interest. One of the indicators that is considered very important in influencing consumer buying interest and trust is OCR. This study aims to analyze OCR clustering in one of the marketplaces in Indonesia using the K-Means Clustering Method. K-Means is a clustering algorithm that is quite effective because it has the ability to group large amounts of data and with high speed, the K-Means algorithm partitions data into clusters so that they have the similarity of being in one cluster.

**Keywords-** *Clustering, K-Means Clustering, OCR, Marketplace*

### **I. PENDAHULUAN**

Pada era globalisasi perkembangan teknologi sangat pesat, untuk mempermudah pekerjaan banyak aspek yang bisa dimanfaatkan termasuk arus informasi. Dengan menerapkan teknologi komputer diberbagai bidang, misalnya bisnis, pendidikan, hiburan, kesehatan, kuliner dan dan lan sebagainya. Terlebih pada bidang bisnis saat ini sangat menuntut untuk menggunakan teknologi komputer supaya bisa bersaing dengan kompetitor, yang disebut E-Commerce, Misalnya Bukalapak adalah salah satu situs E-commerce yang merupakan salah satu online marketplace terkemuka di Indonesia. Buka lapak menyediakan sarana jual beli dari konsumen ke konsumen. Siapapun dapat membuka toko online di Bukalapak dan melayani pembeli dari seluruh indonesia untuk transaksi satu maupun banyak dan memberikan jaminan 100 persen uang kembali kepada pembeli jika barang tidak dikirimkan oleh pelapak, atas dasar itulah banyak orang yang berbelanja di Bukalapak seperti harga lebih murah serta pelayanan yang maksimal, harga yang lebih murah karena para kompetitor bersaing dengan harga, keamanan dalam pengemasan barang dan kecepatan dalam pengiriman barang. Disamping kemudahan dalam berbelanja online banyak resiko menggunakan fasilitas berbelanja online yaitu terkait dengan kualitas barang yang tidak sesuai dengan deskripsi dan hal ini sangat berpengaruh terhadap kepercayaan pembeli untuk melakukan repeat order atau pembelian kembali.

Menurut [1] survey yang dilakukan oleh APJII pengguna internet dari tahun ketahun meningkat, untuk hasil survey tahun 2019-2020 pengguna internet di Indonesia sebesar 196.71 juta jiwa dari total populasi penduduk indonesia 266.91 juta jiwa. Dengan meningkatnya penggunaan internet, khususnya penggunaan E-commerce mendorong para pebisnis untuk menggunakan layanan dan bergabung dengan Bukalapak, konsumen bisa berinteraksi langsung dengan penjual, melakukan tawar menawar, memberikan saran, ataupun memberikan kesan ataupun feedback. Dengan adanya permasalahan- permasalahan diatas dilakukanlah beberapa cara untuk mencegah atau menghindari kekecewaan dari si pelanggan yaitu dengan membuat fitur yang terdapat pada toko

online yaitu Online Customer Review (OCR). OCR dianggap merupakan salah satu faktor yang cukup penting dalam memengaruhi kepercayaan maupun minat pembelian pelanggan [2] Online Customer Review merupakan bagian dari Electronic Word of Mouth (eWOM) [3], yaitu merupakan pendapat langsung dari seseorang dan bukan sebuah iklan. Review adalah salah satu dari beberapa faktor yang menentuka keputusan pembelian seseorang [2].eWOM menjadi bentuk advertising yang paling mempengaruhi dalam bentuk komunikasi [4]. Online review menjadi salah satu alat promosi yang sangat berdampak bagi pemasaran, dengan Online review ini akan memberikan pengaruh antara pelanggan untuk mempengaruhi pembelian kembali, dengan adanya informasi ini konsumen bisa mendapatkan kualitas dari produk yang dicari dari ulasan dan pengalaman yang ditulis oleh konsumen yang telah membeli produk dari penjual online [5]. Konsumen biasanya mencari informasi yang berkualitas saat memutuskan membeli produk. Untuk meminimalisasi dampak negatif sebelum melakukan pemesanan barang harus mencari informasi sebanyak mungkin tentang produk tersebut yaitu dengan melihat feedback yang diberikan pelanggan yaitu yang disebut OCR.

### **II. KAJIAN PUSTAKA**

#### **Data Mining**

Data Mining merupakan proses yang menerapkan teknik matematika, statistik, machine learning dan kecerdasan buatan untuk mengidentifikasi informasi yang bermanfaat yang diperoleh dari database besar [6]. Data mining adalah serangkaian proses yang dilakukan untuk menggali nilai tambah yang selama ini tidak diketahui. Tujuan data mining adalah untuk memperoleh pengetahuan yang masih tersembunyi dalam bongkahan data dan pengenalan pola dalam bongkahan data yang akan digali yang disebut pattern recognition.

**K-Mean Clustering**

Salah satu metode partisi atau biasa disebut juga dengan metode non hirarki. Metode *K-Means Clustering* dapat diterapkan pada kasus dengan jumlah objek yang sangat besar.[7]. Dari beberapa teknik *clustering* yang paling sederhana dan umum dikenal adalah *clustering k-means*. Dalam teknik ini kita ingin mengelompokkan obyek ke dalam k kelompok atau *cluster*. Untuk melakukan *clustering* ini, nilai k harus ditentukan terlebih dahulu. Biasanya user atau pemakai sudah mempunyai informasi awal tentang obyek yang sedang dipelajari, termasuk berapa jumlah *cluster* yang paling tepat. Secara detail kita bisa menggunakan ukuran ketidakmiripan untuk mengelompokkan obyek kita. Jika jarak dua obyek atau data titik cukup dekat, maka dua obyek itu mirip. Semakin dekat berarti semakin tinggi kemiripannya. Semakin tinggi nilai jarak, semakin tinggi ketidakmiripannya.[8]

Langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut.[7]

1. Tentukan K atau jumlah cluster
2. Pusat-pusat cluster diberi nilai awal dengan angka random. Tetapi dalam memberikan inisialisasi k pusat cluster dapat dilakukan dengan berbagai cara, dan yang paling serind adalah secara random.
3. Dalam menempatkan objek berada di cluster mana ditempatkan, maka dihitung berdasarkan jarak antara dua objek, demikian juga dengan jarak antara pusat cluster dengan objek. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak euclidean yang dirumuskan sebagai berikut:  
  

$$D(I,j) = X1i - X1j + X2i - X2j + \dots + (Xki - Xkj)^2$$
 Dimana:  
 $D(i, j)$  =Jarak data ke i ke pusat *cluster* j  
 $xki$  =Data ke i pada atribut data ke  
 $Xkj$  =Titik pusat ke j pada atribut ke k
4. Menghitung kembali jarak keanggotaan cluster sekarang dengan pusat cluster, rata-rata dari semua data/objek dalam cluster tertentu merupakan pusat cluster, dengan menggunakan mean (nilai rata-rata) atau median
5. Lakukan kembali pada setiap objek dengan cluster baru, jika pusat cluster baru tidak berganti atau berubah, maka proses clustering berhenti, jika tidak maka akan kembali ke ke tahap ketiga sampai pusat cluster tidak berubah

**Clustering**

Clustering adalah salah satu bagian dari teknik Data Mining. Clustering adalah pengelompokkan sejumlah data atau objek ke dalam cluster (group) sehingga data pada setiap cluster tersebut akan berisi data yang sangat mirip dan berbeda dengan objek dalam cluster lainnya. [8]

Ada dua pendekatan dalam clustering yaitu hirarki dan partisi. Dalam cluster hirarki dimulai dengan membuat cluster dimana setiap cluster beranggotakan satu objek dan berakhir dengan satu cluster dimana anggota m objek, pada setiap tahap prosedurnyam satu cluster digabung dengan satu cluster lainnya, lalu dipilih cluster yang diinginkan dengan menentukan cut off tingkat tertentu, sedangkan dalam partisi pengelompokan objek dimasukkan ke dalam k cluster, dapat dilakukan dengan menentukan pusat cluter awal lalu dilakukan relokasi objek berdasarkan kriteria tertentu sampai dicapai pengelompokan yang optimum.

**Online Customer Reviews (OCR)**

Online Customer Reviews adalah feedback atau ulasan yang diberikan oleh konsumen terkait dengan informasi dari evaluasi suatu produk tentang berbagai macam aspek, dengan adanya informasi tersebut, konsumen bisa memperoleh informasi dari

kualitas dari suatu produk yang dicari dari ulasan dan pengalaman yang ditulis oleh konsumen yang telah membeli produk tersebut dari penjualan online.[9]

Review adalah bagian dari Electronic Word of Mouth, yaitu merupakan pendapat langsung dari seseorang dan bukan sebuah iklan [3]. Review merupakan salah satu faktor yang sangat mempengaruhi yang menentukan keputusan seorang pembeli. Oline Customer Reviews menjadi alat promosi yang sangat baik untuk komunikasi pemasaran[10]. Masyarakat lebih tertarik promosi secara Word of Mouth dibandingkan iklan, karena dianggap informasi tersebut datang dari kalangan yang sudah menggunakan produk tersebut sebelumnya [11]. Dengan adanya review yang diberikan oleh konsumen memberikan efek pada persepsi kredibilitas.

**III. METODE PENELITIAN**

Untuk melakukan proses clustering dalam penelitian ini penulis melakukan analisis Online Customers Reviews (OCR) yang bersumber dari pada kolom komentar yang disediakan pada situs jual beli online. Data OCR yang digunakana untuk pengujian model penelitian diperoleh dari situs jual beli online Bukalapak.com

Tabel 1. Data Sumber

No	Produk	Jumlah Ulasan	Sumber
1	Gamis /Pakaian Wania Syfarosw Syari	887	www.Bukalapak.com

Dalam menentukan review customer apakah respon negatif atau positif ditentukan berdasarkan komentar yang diberikan konsumen terhadap produk yang terdapat pada kolom komentar.

Tabel 2. Model Reviews

Jenis Respon	Komentar
Respon negatif	Cacat, buruk, hancur, jelek, lambat, kecewa,kesal, sobek,rusak,tipu
Respon positif	Baik, keren, awet, bagus, baik, cantik,cepat, lembut, profesional, sesuai, ramah

**PENGUMPULAN DATA**

Pengumpulan data (data crawling) digunakan untuk menjaring reviews yang diberikan customer. Penelitian ini menggunakan data ulasan produk yang diperoleh dari salah satu situs jual beli online. Data dikumpulkan menggunakan aplikasi Octoparse yaitu sebuah aplikasi open source untuk web crawler.

**TEXT PROCESSING**

Pada pemrosesan awal atau text procesing merupakan tahapan kedua pada text mining. Pada tahapan pemrosesan awal bertujuan untuk mempersiapkan data untuk dapat dipakai pada tahan penemuan pola, misalnya mengeliminasi data yang mengandung noise, data tidak konsisten dan data kurang lengkap.

**SELEKSI FITUR**

Pada tahap ini terdapat dua tahapan yang dilakukan, yaitu sebagai berikut:

1. Case Folding  
Case folding merupakan proses mengubah huruf besar menjadi huruf kecil semua.

**Contoh:**

Dia berMain LompaT tAli  
SAYA MAU MAKAN

**Menjadi:**

dia bermain lompat tali  
saya mau makan

2. Non Alphanumeric Removal  
Non Alphanumeric Removal adalah tahapan dimana menghilangkan tanda baca angka.

**Contoh:**

Saya ketiduran...semalam  
TOLONG DISELESAIKAN!!!! KAPAN SIAPNYA  
JIKA SEPERTI INI????!

**Menjadi:**

Saya ketiduran semalam  
Tolong diselesaikan kapan siapnya jika seperti ini

3. Own Stop Words Removal  
Proses Own Stop Words Removal yaitu menghilangkan atau ekspresi

**Contoh:**

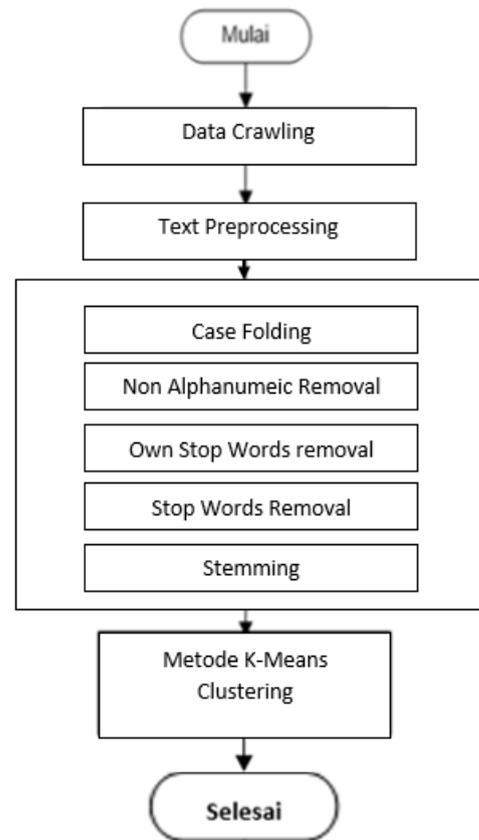
harga bawang merah Rp. 20.000,00  
listrik koq padam, warga jadi bingung mau  
ngapain#ediscirhat  
telah diblokir website <http://www.humor.com>

**Menjadi**

harga bawang merah Rp 20.000,00  
listrik padam, warga sedih  
telah blokir website

4. Stop Words Removal  
Pada tahapan Stop Words Removal dilakukan tahapan dimana kata umum yang sering muncul dalam jumlah yang banyak tetapi tidak memiliki arti. Seperti “ke”, “jika”, “yang”, “akan”, “di”, dan lain sebagainya. Untuk itu perlu dilakukan penghapusan. Untuk melakukan penghapusan kata ini diperlukan sebuah data atau daftar kata yang diinginkan untuk dihapus.

5. Stemming  
Setelah semua data di transformasikan ke dalam bentuk angka, maka data telah dapat dikelompokkan dengan metode K-Means Clustering. Untuk dapat melakukan pengelompokan data tersebut menjadi beberapa cluster, perlu dilakukan beberapa tahapan yaitu:  
1) Tahap awal tentukan terlebih dahulu jumlah cluster yang diinginkan. Dalam penelitian ini data akan dikelompokkan menjadi 2 cluster.  
2) Tentukan titik pusat awal cluster. Dalam penelitian ini titik pusat awal dibangkitkan secara acak atau random



Gambar 1. Gambaran Umum Sistem

**IV. ANALISA DAN HASIL**

**1. Hasil pengumpulan data**

Penelitian ini menggunakan data customer online review sebanyak 887 data.

<ol style="list-style-type: none"> <li>1. Barang sudah diterima dengan kondisi baik</li> <li>2. bagus...!</li> <li>3. Barang SESUAI gambar GAESS</li> <li>4. CEPAT</li> <li>5. SEPATUNYA bagus...!</li> <li>6. Sesuai YANG ditawarkan semoga awet</li> <li>7. bagus dan cepat sampe :P</li> <li>8. parahhhhhhhh...\$\$\$@#@@:(</li> <li>9. TERIMAKASIH bukalapak...</li> <li>10. KECEWA ☹️</li> <li>11. bagus kualitas!!!!</li> <li>12. Barang bagus selanjutnya pesan lagi @.@</li> <li>13. BAGUS 😊</li> <li>14. Puas...</li> <li>15. Sepatunya Bagusss dan anak saya suka terimakasih</li> </ol>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Gambar 2. Data OCR

**2. Pre Processing Data**

Setelah data OCR telah dicrawling, maka proses selanjutnya dilakukan pre processing agar data OCR dapat diterapkan pada algoritma clustering. Tahapan pre processing yang diterapkan adalah case folding, Non Alpha Numeric, Stop Words Removal dan Stemming. Daftar stop words untuk bahasa Indonesia terdiri atas 760 kata.



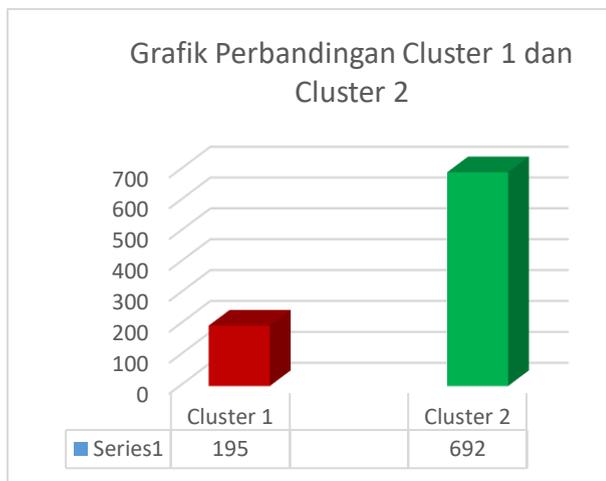
**9. Cluster Instance**

Tabel 4. Cluster Instance

Cluster 1	2	22%
Cluster 2	7	78%

Dari tabel diatas disimpulkan bahwa hasil pengujian terhadap data 887 data OCR menghasilkan 2 cluster yaitu:

- 1) Cluster 1 yaitu menghasilkan sebanyak 22 % (195) ulasan yang mempunyai kemiripan yang sangat tinggi yang dikelompokkan menjadi 1 cluster, yaitu cluster 1
- 2) Cluster 2 menghasilkan sebanyak 78% (692) ulasan yang mempunyai kemiripan yang sangat tinggi yang dikelompokkan menjadi 1 kelompok cluster yaitu cluster 2.



Gambar 9. Grafik Perbandingan Cluster 1 dan Cluster 2

**V. KESIMPULAN**

Dari penelitian yang telah dilakukan, maka diambil kesimpulan sebagai berikut:

- 1) Data Online Customers Reviews adalah 887, dengan 15 atribut.
- 2) Jumlah cluster sebanyak 2
- 3) Dari seluruh data OCR sebanyak 887 data, diperoleh Cluster 1 yaitu sebanyak 195 reviews yang memberikan respon negatif, dan Cluster 2 yaitu sebanyak 692 reviews yang memberikan respon positif.
- 4) Hasil evaluasi cluster dari data Online Customer Review (OCR) pada online marketplace, dapat membantu dan sebagai bahan pertimbangan calon pembeli dalam menentukan keputusan dalam pembelian sebuah produk.

**DAFTAR PUSTAKA**

[1] Asosiasi Penyelenggara Jasa Internet Indonesia, "Laporan Survei Internet APJII 2019 – 2020," *Asos. Penyelenggara Jasa Internet Indones.*, vol. 2020, pp. 1–146, 2020, [Online]. Available:

<https://apjii.or.id/survei>.

[2] A. N. Ardianti and M. A. Widiartanto, "Pengaruh Online Customer Review dan Online Customer Rating terhadap Keputusan Pembelian melalui Marketplace Shopee.," *J. Ilmu Adm. Bisnis*, pp. 1–11, 2019.

[3] S. T. Adeliastari, Vina Ivana, "ELECTRONIC WORD-OF-MOUTH (e-WOM) DAN PENGARUHNYA TERHADAP KEPUTUSAN PEMBELIAN DI RESTORAN DAN KAFE DI SURABAYA Adeliastari," no. 2010, pp. 218–230, 2017.

[4] A. A. Laksmi and F. Oktafani, "PENGARUH ELECTRONIC WORD OF MOUTH (eWOM) TERHADAP MINAT BELI FOLLOWERS INSTAGRAM PADA WARUNG UPNORMAL," *J. Comput. Bisnis*, vol. 10, no. 2, pp. 78–88, 2016, [Online]. Available: [www.dailysocial.id](http://www.dailysocial.id).

[5] A. Farki, I. Baihaqi, and M. Wibawa, "Pengaruh online customer review rating terhadap kepercayaan place di indonesia," vol. 5, no. 2, 2016.

[6] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

[7] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.

[8] L. M. Pratiwi, Diana, and E. P. Agustin, "Penerapan K-Means Clustering Untuk Memprediksi Minat Nasabah Pada Pt . Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih," *Univ. Bina Darma*, pp. 1–16, 2016.

[9] N. R. Ayu and D. L. Chaerowati, "Hubungan Online Customer Review pada Media Sosial Instagram," pp. 410–417.

[10] *et al.*, "Evaluasi Cluster Social Media Data In Tourism Domain Menggunakan K-Means Clustering," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 4, no. 1, pp. 89–93, 2020, doi: 10.46880/jmika.v4i1.148.

[11] S. Natania, "Analysis Of The Effect Of Online Customer Review On Millennial’s Purchase Decision In Bandung (Case Study Of Gadget Products At Tokopedia)," no. 227, pp. 1–17, 2018.