JAIIT (Journal of Advances in Information and Industrial Technology)
Vol. 3, No. 2 November 2021, ISSN 2723-4371, E-ISSN 2723-5912

Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python

Ainur Rahman[1)], Heri Suroyo[2)]
[1, 2)]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma Palembang,
Jalan Jendral Ahmad Yani No.3, Palembang, 30111, Indonesia.
Email: ainurrahman.bae@gmail.com[1)], herisuroyo@binadarma.ac.id[2)]

Abstrak

Fokus dari peneltitian ini adalah melakukan analisis *text mining* pada produk *elektronik* yang dijual di *e-commerce* Shopee dengan menggunakan metode *Kmeans algoritm* memakai *python* sebagai bahasa pemrograman. Data di *scraping* berupa teks komentar, angka penjualan dan skor rating bintang. Data hasil dari penelitian didapatkan pada analisis teks komentar produk dengan *wordcloud* produk *Smartphone low cost* menunjukan data komentar *marketplace* shopee Indonesia dapat bahwa baik di *smartphone low cost* maupun yang *high cost* cenderung memiliki pola *wordloud* yang sama dimana kata-kata yang dominan muncul cenderung *netral* dan positif, sedang kata-kata yang bermakna negatif cederung tidak dominan. Sementara kata yang sering muncul yaitu barang, mantap, cepat,kirim dan bagus dengan nilai akurasi presentase sebesar 92%. Sedangkan hasil proses *wordcloud medium high cost* diperoleh kata yang sering muncul ialah kata (kirim, cepat, dan bagus) dengan nilai akurasi dengan presentase 94%. Serta berdasarkan hasil grafik dari proses *clustering* data *k-means* menunjukan bahwa angka penjualan 0 sampai 1000 mendapatkan skor rating bintang tertinggi dan penjualan dengan skor rating bintang terendah ialah antara 1500 sampai 2000 ke atas.
Kata kunci: *python, scraping, clustering.*

1. Pendahuluan (Introduction)

Perkembangan teknologi informasi diera *modernnisasi* yang secara tidak langsung menuntut melakukan sesuatu dengan serba cepat dan menimbulkan dampak yang besar dan signifikan dalam kehidupan sehari-hari manusia salah satunya ialah terciptanya teknologi *internet.* Adapun kegiantan yang dapat dilakukan dengan pemanfaattan teknologi internet ialah melakukan proses bisnis penjualan atau pembelian berbasis *online* atau biasa di sebut *e-commerce* (Jony Wong, 2008).

Untuk mengetahui tingkat kepuasan dari setiap pelanggan terhadap kualitas produk yang dijual. Setiap *e-commerce* menciptakan atau membuat sebuah kolom *review* atau komentar yang bertujuan sebagai tempat atau wadah bagi pelanggan untuk menyampaikan dan membemberian ulasan mengenai kualitas produk yang di jual. Dimana dari setiap ulasan yang di berikan pelanggan tersebut nantinya memiliki pengaruh terhadap penjualan di *e-commerce* tersebut, karna setiap pelanggan yang ingin membeli sebuah produk tentunya melihat terlebih dahulu ulasan mengenai kualitas produk tersebut.

Beberapa penelitian yang relevan diantaranya text analis cluster akun shopee Indonesia memakai komentar pengguna memakai *orange data mining* (Sentiya et al., n.d.) dan penerapan cluster analis serta sentimen data twiter pada penilaian wisata pantai memakai *Kmeans method* (Syaifudin & Irawan, 2018). Penelitian ini mencoba menganalisis ulasan atau komentar dari dua *smartphone medium quality* yaitu beberapa smartphone produk china dengan kualitas yang sama serta harga yang sama pada *e-commerce* shopee Indonesia. Adapun hasil dari dilakukannya penelitian ini adalah analisis isu pengelompokkan istilah yang lebih banyak didominasi. Dalam melakukan penelitian ini penulis menggunakan teknik analisis teks *clustering* dengan metode *k-means*. Adapun bahasa pemrograman yang digunakan dalam melakukan penelitian ini ialah *python*. Serta dalam proses pengumpulan data menggunakan metode *scraping* yang berarti pengambilan sebagian data atau konten dari suatu situs *web*. Dengan memafaatkan aplikasi *webscraping* (Indraloka & Santosa, 2017).

2. Metode Penelitian (Methods)

Analisis teks pada penelitian dilakukan dengan menarik (*scrapping*) data komentar dari produk yang menjadi objek penelitian di *e-commerce*. Text analysis artinya menyiratkan bahwa itu adalah kesalahan dari salah satu bagian awal semiotika teks, yang secara eksplisit berbicara tentang teks sebagai hasil dari penggunaan bahasa yang merupakan campuran atau seperangkat tertentu atau secara eksplisit menyangkut kerangka sintaksis atau tanda paradigmatik. lapisan yang menunjukkan indikasi atau implikasi. hubungan antara tanda-tanda figuratif atau metonimik, konten fantasi, dan filosofi di baliknya. Karena penyelidikan teks dan semiotika teks adalah bagian dari semiotika umum, aturan-aturan penting yang membentuk semiotika luas juga berlaku untuk mereka. yang memiliki arti penting, meskipun unit kajian yang paling kecil adalah semiotika yang mengandung makna teks, namun teks tidak dapat dipisahkan dari tanda-tanda yang menyusunnya (Piliang, 2004).

Sementara analisis *clustering* dilakuan pada penelitian dengan mengambil data angka penjualan dan tingkat rating bintang yang diberikan *customer*. Strategi k-implies/kmeans merupakan teknik pemeriksaan informasi atau information mining strategy yang memainkan sistem demonstrasi tanpa manajemen (solo) dan merupakan salah satu teknik yang melakukan pengumpulan informasi dengan kerangka parsel. Strategi k-implies bunching mencoba untuk mengelompokkan informasi terkini ke dalam beberapa gathering, dimana informasi dalam satu gathering memiliki atribut yang sama satu sama lain dan memiliki berbagai kualitas dari informasi pada gathering yang berbeda.Adapun tahapan-tahapan dalam metode *k-means* adalah:

1. Tentukan jumlah kelompok yang perlu Anda buat.
2. Perkenalkan k sebagai centroid yang dapat dibuat sewenang-wenang.
3. Hitung jarak setiap informasi ke setiap centroid dengan menggunakan kondisi Euclidean Distance.
4. Mengumpulkan setiap informasi tergantung pada jarak terpendek antara informasi dan pusatnya.
5. Tentukan situasi centroid baru (k).
6. Kembali ke tahap 3 jika situasi centroid baru dengan centroid lama tidak terlalu mirip.

3. Hasil dan Pembahasan (Results and Discussions)

Data diambil atau *scrap* dari *e-commerce* shopee.co.id selanjutnya akan dilakukan *pre-prosesesing* antara lain: *cleaning, remove stopword, tokenization steming.*

3.1 *Pre-Prosesesing*

a. Cleaning

Ialah proses penghapusan tanda baca(*punctuation*), *symbol,* huruf besar (*uppercase*) manjadi huruf kecil (*lowercase*) serta bilangan angka(*numbers*). Sehingga data tersebut menjadi data yang *efektif* serta dapat diolah dengan baik.

b. Remove Stopword

Ialah sebuah proses menghapus (kata-kata) yang kurang penting atau *words* yang tidak penting misalnya kata dan, atau, kamu, aku.

c. Tokenization

Ialah Metode yang terlibat dengan memecah kalimat menjadi beberapa bagian dikenal sebagai token. Token dapat dianggap sebagai jenis kata, ekspresi, atau komponen yang memiliki makna.

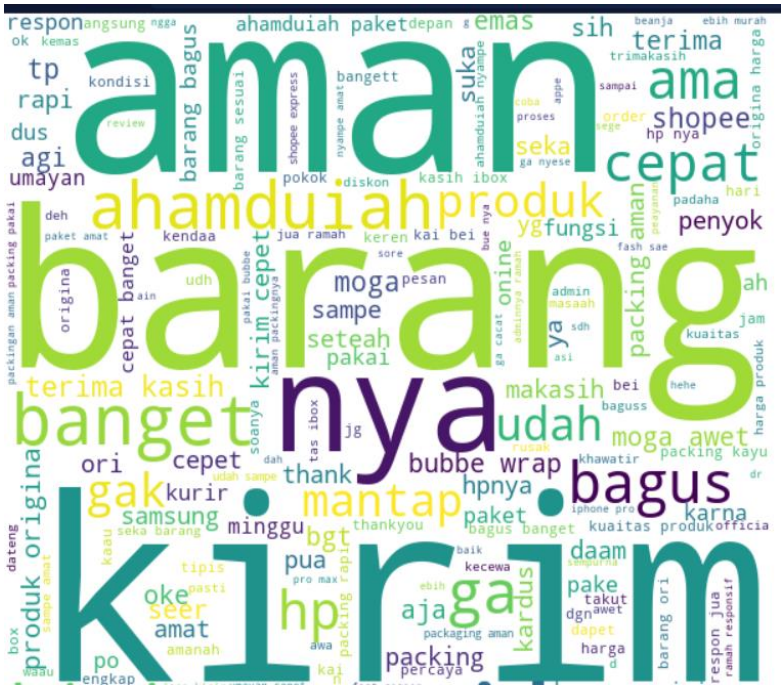
d. Stemming

Ialah cara paling umum untuk mengubah kata ke dalam struktur dasarnya dengan menghilangkan gabungan sebelumnya, kemudian setelah kata itu.

 DOI: https://doi.org/10.52435/jaiit.v3i2.158

3.2 Hasil dan analisis *Word cloud*

Word cloud ataupun biasa disebut juga dengan *text cloud* merupakan satu dari beberapa metode untuk menampilkan data teks secara visual. Metode ini sangat banyak digunakan dan popular terutama untuk teks minning. Berikut ialah hasil dari proses analisis teks *word cloud.*



Gambar 3.1 *Output* dari proses *word cloud smartphone jenis low cost*

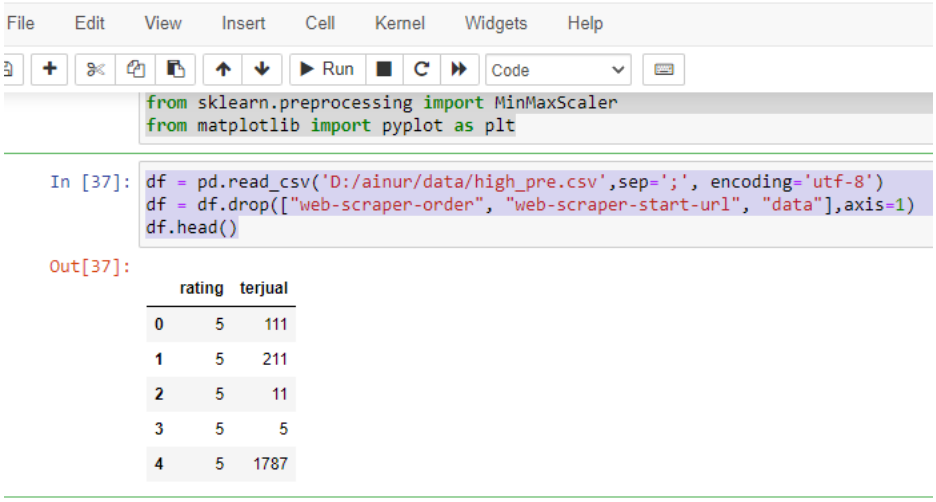
Dari *Word cloud graphics* diatas nampak kata dominan adalah aman, barang dan kirim, beberapa kata yang lebih kecil muncul antara lain cepat, bagus, cepet, penyok, dan terima. Nampak juga beberapa kata yang bermakna negatif seperti penyok, takut, dan setengah pakai hanya muncul dengan format huruf yang kecil yang berarti tidak mendominasi teks. Sedang kata-kata yang dominan cenderung netral atau positif seperti barang, kirim, aman dan bagus.

Gambar 3.2 Hasil atau *output* dari proses *word cloud smartphone jenis high cost*

Berdasarkan hasil proses *word cloud smartphone jenis high cost* pada gambar di atas menunjukan bahwa kata yang dominan sering muncul pada ialah kata barang, mantap, cepat,kirim dan bagus. Sebagian cenderung kata-kata netral dan positif. Sementara kata-kata yang bermakna positif banyak muncul namun tidak dominan misalnya kata kirim, cepat, ramah, bagus banyak muncul di format teks yang kecil artina tidak terlalu mendominasi komentar. Sedangkan kata yang negatif hampir tidak nampak pada *wordcloud*.

3.3 *K-Means Clustering*

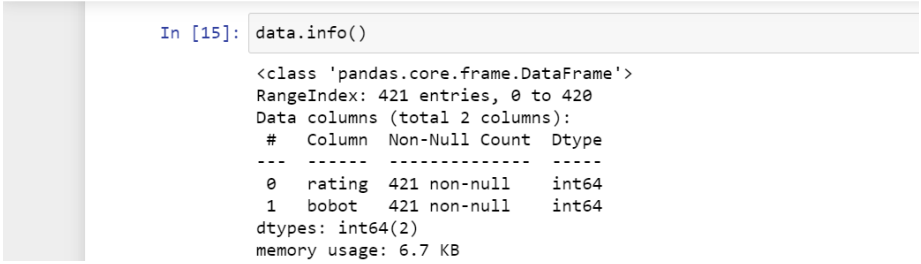
Analisis *clustering* dilakukan pada data penjualan dan rating yang diperoleh pada setiap item barang smartphone yang ditarik dari *e-commerce*. Berikut proses dan hasil analisis *K-Means* dengan *python* untuk data rating dan angka penjualan pada *smartphone low cost*.



Gambar 3.3 *Code import library* dan hasil atau *output* data

Dari data di atas untuk *smartphone low cost* nampak bahwa angka penjualan baik yang rendah maupun yang tertinggi cenderung memiliki *cluster* yang sama dengan kecenderungan sama memperoleh rating 5 dari pelanggan.

#data info
Data.info()



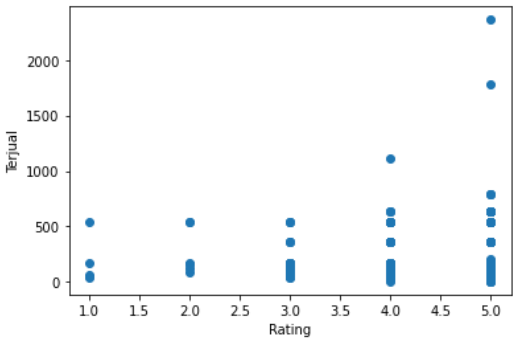
Gambar 3.4 Melihat informasi data atau data info

Sebaran data
#sebaran data
plt.scatter(df.rating,df['terjual'])
plt.xlabel('Rating')
plt.ylabel('Terjual')

DOI: https://doi.org/10.52435/jaiit.v3i2.158

```
In [38]: #sebaran data
         plt.scatter(df.rating,df['terjual'])
         plt.xlabel('Rating')
         plt.ylabel('Terjual')

Out[38]: Text(0, 0.5, 'Terjual')
```



Gambar 3.5 Proses melihat sebaran dan hasil *(output)*

***k-means* prediksi**

df['cluster']=y_predicted

df.head(10)

	rating	terjual	cluster
0	5	111	1
1	5	211	1
2	5	11	1
3	5	5	1
4	5	1787	2

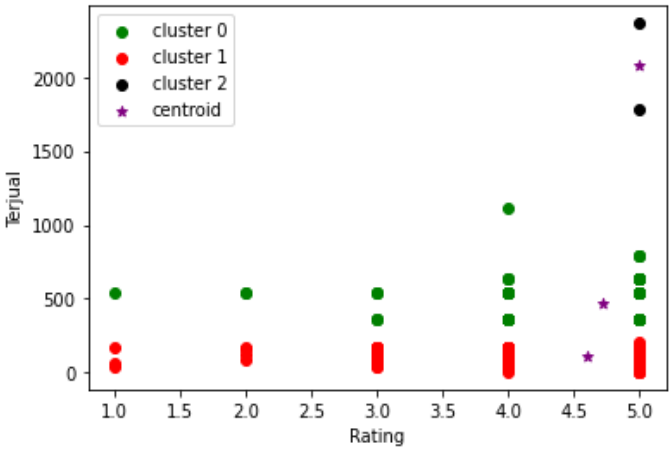
Gambar 3.6 hasil prediksi

Dari data di atas untuk smartphone *high cost* juga nampak tidak ada perbedaan dimana angka penjualan baik yang rendah maupun yang tertinggi cenderung memiliki *cluster* yang sama dengan kecenderungan sama memperoleh rating 5 dari pelanggan, meskipun terdapat *cluster* yang berbeda untuk angka penjualan yang tertinggi.

Selanjutnya analisis *clustering* untuk data yang di *scrap* baik data *smartphone* yang *high cost* maupun yang *low cost* diperoleh hasil sebagai berikut.

Visualisai *clustering.*

```
Out[42]: <matplotlib.legend.Legend at 0x18e9b8dd100>
```



Gambar 3.7 proses dan hasil *output 3 cluster data seluruh smartphone*

Berdasarkan hasil grafik dari proses *clustering* di atas menunjukan bahwa angka penjualan 0 sampai 1000 mendapatkan rating tertinggi.

Keterangan:

1. Warna bintang ungu berarti menunjukan titik pusat atau lokasi *cluster* yang biasa di sebut juga *centroid.*
2. Warna hijau merupakan *cluster* pertama atau *cluster* 0 dikarenakan di *python* angka dimulai dari 0 dan murupakan *cluster* dengan nilai sedang.
3. Warna merah merupakan *cluster* kedua atau *cluster* 1 dan merupakan *cluster* dengan nilai tertinggi.
4. Warna hitam *cluster* ketiga atau *claster 2* dan merupakan *cluster* dengan nilai terendah.

4. Kesimpulan (Conclusion)

Berdasarkan uraian pada bab sebelumnya serta dari hasil proses analisis pengelompokan teks selesai pada keterangan pusat niaga shoppe Indonesia, didapatkan beberapa kesimpulan:

1. Berdasarkan data komentar *marketplace* shopee Indonesia dapat bahwa baik di *smartphone low cost* maupun yang *high cost* cenderung memiliki pola *wordloud* yang sama dimana kata-kata yang dominan muncul cenderung netral dan positif, sedang kata-kata yang bermakna negatif cederung tidak dominan.
2. Hasil analisis *clustering* cenderung diperoleh tidak ada beda antara angka penjualan dengan rating yang diperoleh baik untuk produk *smartphone high level* maupun *low level*.
3. Dari proses *clusterisasi* menggunakan *k-means* terdapat 3 *cluster* yaitu *cluster* pertama(0) ditunjukan dengan warna hijau, *cluster* kedua(1) dengan warna merah serta *cluster* ketiga(2) dengan warna hitam . Warna bintang ungu berarti menunjukan titik pusat *cluster* atau menunjukan titik *centroid* nya.

Ucapan Terima Kasih (Acknowledgement)

Penulis menyampaikan terima kasih sebesar-besarnya kepada seluruh tim atas kerja bersamanya, khususnya untuk Bapak Heri Suroyo, beliau merupakan dosen pembimbing kami yang telah memberikan saran, bimbingan serta referensi untuk publikasi di jurnal ini.

Daftar Pustaka

A. Yani,D.D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. Jurnal Sistem Dan Teknologi Informasi (JUSTIN), 7(4), 257. https://doi.org/10.26418/justin.v7i4.30930

 DOI: https://doi.org/10.52435/jaiit.v3i2.158

Handoyo,R., Rumani,R., & Nasution,S.M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan KMeans Pada Pengelompokan Dokumen. JSM STMIK Mikroskil, 15(2), 73–82. https://mikroskil.ac.id/ejurnal/index.php/jsm/article/view/161

Sentiya, A., Suroyo, H., Komputer, F. I., & Darma, U. B. (n.d.). bina darma conference on computer science ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE INDONESIA DENGAN bina darma conference on computer science. 1055–1067.

Hasibuan A, Z. (2007). Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi, *4*(1), 1–194.

Jony Wong, W. (2008). Belanja Elektronik. Belanja Elektronik, 10–42.

Syaifudin, Y. W., & Irawan, R. A. (2018). Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means. Jurnal Informatika Polinema, *4*(3), 189. https://doi.org/10.33795/jip.v4i3.205

Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. Jurnal Sains Dan Seni ITS, 6(2), 6–11. https://doi.org/10.12962/j23373520.v6i2.24419

Halaman ini sengaja dikosongkan

 DOI: https://doi.org/10.52435/jaiit.v3i2.158