Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X
Vol. 6, No. 1, Januari 2022, hlm. 82-89 http://j-ptiik.ub.ac.id

Pengelompokan Ulasan Produk HP pada Marketplace Tokopedia menggunakan Metode *Semi Supervised K-Means*

Rizky Ardiawan[1], Yuita Arum Sari[2], Bayu Rahayudi[3]

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: [1]rizky19@student.ub.ac.id, [2]yuita@ub.ac.id, [3]ubay1@ub.ac.id

Abstrak

Internet sudah berkembang pesat sesuai dengan perubahan zaman. Hal tersebut juga merubah perilaku berbelanja yang awalnya bertatap muka sekarang bisa dilakukan secara *online*. Hp atau *smartphone* merupakan barang yang paling banyak dicari di zaman sekarang. Untuk membeli barang tersebut secara *online* banyak sekali *marketplace* yang tersedia di Indonesia seperti Tokopedia. Sebuah ulasan produk dinilai sebagai faktor utama untuk konsumen membeli barang. Untuk melakukan analisis pada ulasan diperlukan metode yang dapat mengklasifikasikan dan mengelompokkan ulasan terhadap kategori yang ada.Dengan menggabungkan dua pemahaman antara *Supervised* dan *unsupervised* dapat menciptakan sebuah metode pengelompokan yang berdasarkan data latih yang terdiri data yang berlabel. Metode yang cocok untuk kasus tersebut adalah metode *Semi Supervised K-Means*. Dari hasil penelitian ini bahwa dalam 4 percobaan yang berbeda didapatkan evaluasi nilai klaster dengan menggunakan *Silhouette* sebesar 0,013647 yang merupakan nilai terbesar dengan menggunakan metode *Semi Supervised K-Means*. Nilai *Silhouette* yang dihasilkan oleh penelitian ini sangat kecil dikarenakan jumlah dimensi yang sangat besar untuk mengelompokan *cluster* yang sangat sedikit yaitu sebesar 3 *cluster*. Akan tetapi hasil pengelompokan *cluster* yang dihasilkan pada metode yang sama terbukti lebih baik dari pada metode *K-Means* pada umumnya dengan data ulasan sesuai pada label aslinya.

Kata kunci: Hp, Ulasan, *Marketplace*, Tokopedia, Pengelompokan Ulasan, *Semi Supervised K-Means*

Abstract

The internet has grown rapidly in accordance with the changing times. It also changes shopping behavior that was originally face to face now can be done online. Cell phones or smartphones are the most sought after items today. To buy these items online, there are many marketplaces available in Indonesia, such as Tokopedia. A product review is rated as the main factor for consumers to buy goods. To perform analysis on reviews, a method is needed that can classify and group reviews into existing categories. By combining the two understandings between Supervised and unsupervised, one can create a grouping method based on training data consisting of labeled data. The method that is suitable for this case is the Semi Supervised K-Means method. From the results of this study, it was found that in 4 different experiments, the evaluation of the cluster value using Silhouette was 0.013647 which was the largest value using the Semi Supervised K-Means method. Which is very small, namely 3 clusters. However, the results of clustering the clusters produced in the same method proved to be better than the K-Means method in general with the review data according to the original label.

Keywords: *Mobile, Reviews, Marketplace, Tokopedia, Review Grouping, Semi Supervised K-Means*

1. PENDAHULUAN

Internet sudah menjadi kebutuhan utama di zaman sekarang. Tidak heran banyak masyarakat Indonesia mulai berbondong-bondong untuk masuk kedalam dunia internet untuk mengikuti perkembangan zaman yang ada. Salah satu yang menjadi tujuan masyarakat Indonesia adalah dunia *e-comerse*. Dalam hal ini Masyarakat Indonesia lebih menyukai berbelanja secara *online* dibandingkan secara *offline* (Harahap, 2018).

Tokopedia merupakan *market place* yang berasal asli dari Indonesia ,dan juga yang paling diminati dalam negaranya sendiri. Menurut data yang diperoleh dari Similarweb menyebutkan bahwa Tokopedia menjadi *Market place* yang

menempati urutan teratas dalam *Trafic Market* atau dari total pengunjung. Disebutkan bahwa total kunjungan Tokopedia sebanyak 131,4 juta pengunjung lebih banyak dari kopetitornya.

Market place juga merupakan tempat berbelanja paling efektif dan efisien di zaman sekarang, dikarenakan kita tidak terlalu memakan banyak waktu pergi ke toko untuk membeli barang. Hp(*Handphone*) atau biasa orang menyebut *smartphone* menjadi barang yang keberadaannya menjadi sangat penting dikarenakan di zaman sekarang semua serba *digital* yang artinya semua kegiatan dapat dipantau, diakses atau dilakukan melalui Hp. Karenakan sangat pentingnya keberadaannya beberapa brand Hp ternama di dunia mulai menjajaki pasar Hp yang ada di Indonesia.Terdapat 3 brand teratas menurut data dari IDC yaitu dari brand Vivo untuk Y-series, Oppo untuk Reno Series dan Xiaomi untuk redmi note 9 pro .

Dalam penentuan produk Hp yang akan mereka beli maka dibutuhkan faktor-faktor seperti kepercayaan, harga, kenyamanan, kemudahan dan ketersediaan yang dinilai sebagai faktor utama dalam pemilihan barang(Harahap, 2018). Dari hal tersebut bahwa kepercayaan menjadi faktor utama yang mendasari seseorang untuk membeli barang atau tidak ini bisa dibuktikan di penelitian terdahulu bahwa dimensi trust memiliki pengaruh yang signifikan secara positif terhadap sikap, niat dan perilaku konsumen dalam berbelanja secara online(Assegaff, 2015).

Berangkat dari ulasan tersebut konsumen dapat melihat suatu penilaian terbaru atau parameter terbaru dalam sikap dan niat untuk membeli suatu produk pada *marketplace* Tokopedia. Dalam teks *mining* terdapat 2 metode untuk mempermudah dalam melakukan analisis masalah pada ulasan tersebut yaitu *supervised* dan *unsupervised*. Penggabungan antara metode *supervised* dan *unsupervised* diharapkan dapat membantu untuk dapat mengelompokkan sebuah dokumen berdasarkan data latih yang ada dan juga dapat dikelompokkan berdasarkan label atau kategori yang ada

Pernyataan diatas bisa disimpulkan bahwa kepercayaan menjadi faktor utama dalam penentuan pembelian barang atau tidak. Untuk melihat kepercayaan itu bisa dilihat dari hasil *review* yang telah dilakukan oleh konsumen sebelumnya. *Review* menjadi bahan pertimbangan untuk penentuan keputusan akhir tersebut. Hal ini sejalan dengan konsumen akan lebih melihat hasil *review* daripada melihat barang yang ada. Hasil *review* dari konsumen juga berdampak sangat besar dalam penjualan produk daripada peringkat dan jumlah ulasan(Wu et al., 2018).

Dalam penelitian tentang penggunaan metode *K-means dalam* pengelompokan review dari pengguna indosat dinilai berhasil dengan memunculkan 3 buah cluster(Saputra & Arianty, 2019). Dalam pengelompokan teks menggunakan metode *K-Means* penentuan inisiasi awal sebuah cluster sangat berpengaruh dalam mencari keanggotannya(Syaifudin & Irawan, 2018). Oleh karena itu dalam kita harus menentukan cluster yang menjadi titik pusat dari setiap cluster untuk meningkatkan hasil dari pengelompokan teks nantinya(Li et al., 2019). Dalam penelitian tentang *Semi Supervised K-Means* menyebutkan bahwa dalam penelitian tersebut algoritma *Improved Semi Supervised K-Means* menghasilkan pengkelompokan yang efisen terhadap algoritma ini(Gao et al., 2008) .

Dari uraian masalah diatas maka metode *Semi Supervised K-Means* digunakan untuk pengelompokan dari ulasan pembelian produk yan nantinya akan digunakan untuk tahap analisis mengetahaui bahwa produk tersebut apakah produk tersebut layak dibeli atau tidak.

2. Tinjauan Pustaka

2.1 Penelitian Terdahulu

Pada penelitian berikut yang dilakukan oleh (Mahyoub et al., 2019) pada penelitian ini ingin menguji apakah algoritma *Semi Supervised* dapat digunakan kepada data yang belum memiliki stuktur. Data yang diperoleh dari mentranskip video review menggunakan fitur *Speech Recognition*. Sehingga data yang didapatkan murni dari video percakapan tersebut dan memiliki struktur yang tidak teratur. Dalam pengujian metode tersebut peniliti memilih *Silhoutte* untuk mengevaluasi dari hasil pengelompokan data nantinya. Hasil dari penelitian tersebut adalah didapatkan bahwa hasil evaluasi dari metode sebelumnya untuk pengelompokan pada data teks dari video bekerja sangat baik.

2.2 Pengelompokan Dokumen

Pengelompokan dokumen bertujuan untuk mengelompokkan beberapa dokumen yang memiliki kemiripan. Dalam penerapannya

pengelompokan dokumen bertujuan untuk membagi dokumen menjadi beberapa bagian, dari bagian tersebut bisa terlihat kemiripan dari jumlah kata atau term yang mirip.

2.3 *Preprocessing*

Pre Processing Merupakan tahapan dalam teks *mining* yang bertujuan untuk mengolah kata yang masih mentah menjadi kata yang nantinya siap diproses untuk tahapan selanjutnya dari teks *mining* .

2.4 *Case Folding*

Case Folding digunakan untuk mengubah semua semua karakter teks pada dokumen menjadi *lower case* / menjadi huruf kecil. Tujuan dilakukan *case folding* sendiri memberikan nilai yang sama antar kata seperti kata "kebun" dengan "kebun".

2.5 *Stemming*

Stemming merupakan tahapan dalam *pre processing* teks *mining* untuk mengubah sebuah kata kedalam bentuk kata dasarnya. Dilakukannya proses *stemming* untuk menghindari duplikasi kata yang seharusnya memiliki makna atau arti yang sama.

2.6 Tokenize

Tokenize merupakan tahapan yang ada dalam *pre processing* untuk memisahkan atau memotong data menjadi beberapa term/token.

2.7 Stopword Removal

Stopword Removal atau juga dikenal sebagai *filtering* merupakan tahapan dalam *pre processing* digunakan untuk menghilangkan kata/term/token yang dinilai tidak berpengaruh dalam suatu dokumen tersebut.

2.8 Pembobotan *Tf-Idf*

Pembobotan digunakan untuk mengukur seberapa sering term atau kata tersebut muncul dari dokumen tersebut. Menghitung semua kemunculan kata yang muncul lalu dikalikan dengan *Inverse* dari dokumen frekuensinya sehingga rumus (Qaiser & Ali, 2018) dapat dituliskan sebagai persamaan (1) berikut :

$$tf - idf = \left(1 + \log tf_{t,d,}\right) \times \log\left(\frac{N}{dft}\right) (1)$$

Keterangan :

$tf_{t,d,}$ = jumlah frekuensi kemunculan kata/term t dalam dokumen d

N = Banyaknya dokumen

dft = jumlah frekuensi dokumen d yang mengandung term t

2.9 *Semi Supervised K-Means*

Konsep dasar dari *Semi Supervised K-Means* adalah inisiasi jumlah *cluster*, tentukan titik pusat kluster lalu hitung dengan *EcludianDistance*. Dalam hal ini *Semi Supervised K-Means* digunakan untuk mengelompokkan dokumen pada sebuah teks berikut merupakan Langkah-langkah pada metode *semi supervised k-means* sebagai berikut :

1. Tentukan jumlah *cluster* dan Tentukan Titik pusat dari setiap *cluster*
   Tentukan titik pusat dengan mengambil data yang telah berlabel dengan nilai dari *centroid* tersebut diambil dari *normalisasi tf-idf*
2. Hitung *Cosine distance*
   Menghitung *cosine distance* dari pembobotan *tf-idf* yaitu dengan menghitung jarak dari dokumen ke titik pusat *cluster* yang sudah ditentukan dengan menggunakan rumus (2) sebagai berikut :

   $$CosSim(d_j, d_k) = \sum_{i=1}^{t}(W_{ij} \times W_{ik}) (2)$$

   Keterangan :

   W_{ij} = jumlah bobot dari indeks j

   W_{ik} = jumlah bobot dari indeks k

   Setelah menghitung *Cosine Similitary* yang ditunjukan pada persamaan (2) selanjutnya menghitung *Cosine distance* dengan persamaan (3)

   $$Cosine\ distance = 1 - Cosine\ Similitary (3)$$

   Sesudah pada persamaan (3) maka hitung keanggotaan *cluster* dengan mencari nilai minimun dari setiap jarak ke *cluster*,
3. Perubahan nilai *cluster*

$$k_{newi,j} = \sum_{k=0}^{n_i} w_{k,j}/n_i (4)$$

Keterangan pada persamaan (4) :

$k_{newi,j}$ = Nilai rata-rata kluster baru

$w_{k,j}$ = Bobot dari data ke-k yang ada dalam cluster sampai variable j

n_i = Jumlah data sebuah *cluster* ke-i

4. Kondisi Berhenti
   a. Apabila iterasi sudah terpenuhi
   b. Apabila tidak ada perubahan yang terjadi terhadap keanggotaan *cluster* maka kondisi bisa dikatakan konvergen

2.4 *Silhoutte*

Silhouette Coefficient adalah sebuah metode untuk mengukur sebuah kualitas dari sebuah *cluster* yang telah terbentuk. Metode .Berikut Langkah-langkah evaluasi *cluster* menggunakan *Silhouette Coefficient* menurut (Rousseeuw, 1987) sebagai persamaan (5) berikut:

$$S_i = {(a_i - b_i)}\Big/{Max(a_i, b_i)} \quad (5)$$

Keterangan :

S_i = nilai *Silhouette Coefficient*

a_i = nilai rata-rata objek ke-i dengan objek lainnya yang berada pada satu *cluster*

b_i = nilai rata-rata objek ke-i dengan objek lainnya yang berada pada satu *cluster*

Menurut (Rousseeuw, 1987) dapat dituliskan nilai-nilai dari kekuatan *cluster* dengan *silhouette* dapat dilihat pada Tabel 1. seperti dibawah:

Tabel 1. Klasifikasi Nilai *Silhoutte*

No	*Silhoutte*	*Structure*
1	$0{,}75 \leq SC \leq 1$	*Strong Structure*
2	$0{,}5 \leq SC \leq 0{,}7$	*Medium Structure*
3	$0{,}25 \leq SC \leq 0{,}5$	*Weak Structure*
4	$SC \leq 0{,}25$	*No Structure*

2.5 *Kappa Measure*

Kappa Measure adalah metode yang digunakan untuk mengukur kesepakatan antar penilai untuk data yang bersifat Kuantitatif(Kategori). Untuk menghitung *Kappa Measure* dapat digambarkan sebagai rumus menurut (Landis & Koch, 1977) sebagai persamaan (6) berikut :

$$k = \frac{P_0 - P_e}{1 - P_e} \quad (6)$$

Keterangan :

P_0 = Kesepakatan yang diamati meliliki kesamaan penilaian

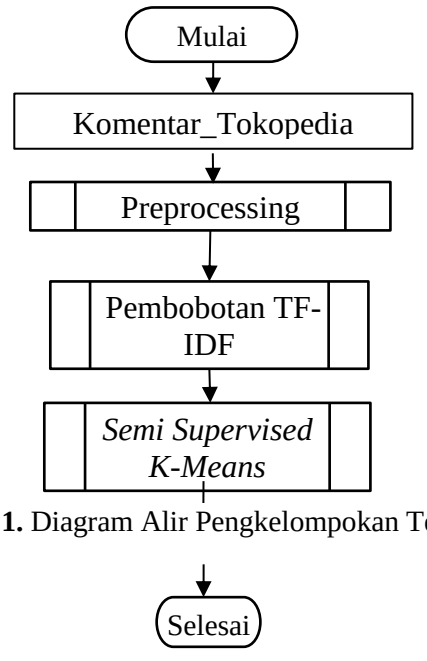
P_e = Kesepakatan yang tidak memiliki kesamaan penilaian

3. METODE PENELITIAN

Dalam penelitian ini, data yang digunakan adalah berupa teks dari ulasan produk hp di tokopedia. Data tersebut diperoleh dengan *crawling* data pada *marketplace* Tokopedia.

3.1 Alur Pengelompokan Teks

Proses ini akan ditunjukkan alur pengkelompokan teks dari proses awal yaitu *Preprocessing* sampai *clustering* menggunakan metode *Semi Supervised* K-Means.



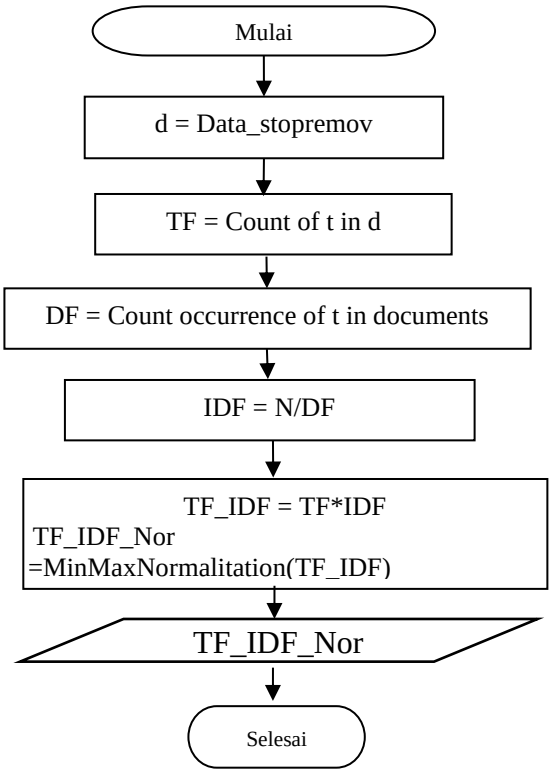
Gambar 1. Diagram Alir Pengkelompokan Teks

Dari Gambar 1.
Seperti diatas maka dapat di artikan sebagai berikut :

1. Pertama Kumpulkan Seluruh Komentar Tokopedia
2. Selanjutnya data Komentar akan di Proses terlebih dahulu dan Dilakukan Pembobotan menggunakan *TF-IDF*
3. Terakhir akan dilakukan *clustering m*enggunakan *semi supervised K-Means*

3.2 Pembobotan TF-IDF

Pada Tahapan ini betujuan untuk memberikan sebuah bobot pada term/kata/token yang nilai dari bobot tersebut akan digunakan untuk pengkelompokan komentar.



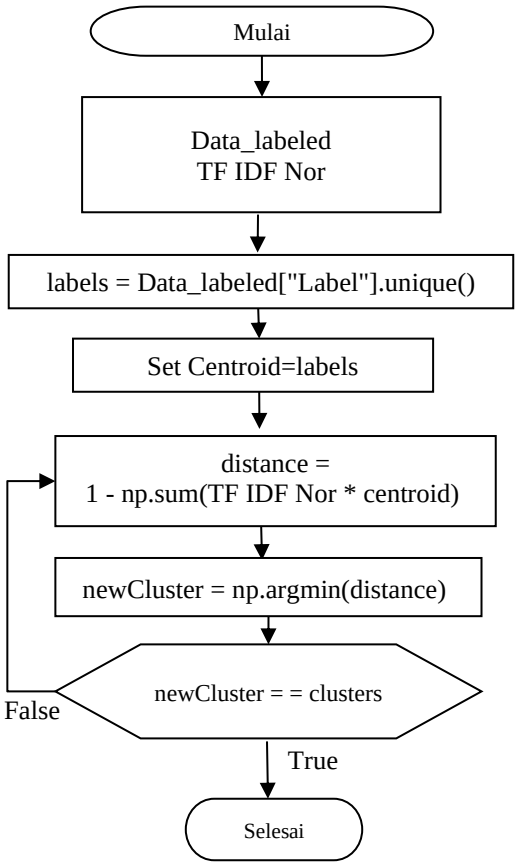
Gambar 2. Diagram Alir Pembobotan TF-IDF

Dari Gambar 2. Seperti diatas maka dapat di artikan sebagai berikut :

1. Pertama Kumpulkan Seluruh Komentar Tokopedia dan telah dilakukan *Preprocessing* data
2. Selanjutnya data Komentar akan di hitung Jumlah dari *Term Frekuensi(TF)* dan *Inverse Document Frekuensi(Idf)*
3. Setelah semua terhitung maka akan dihitung nilai dari *TF-Idf* dengan mengkalikan *Term Frekuensi(TF)* dan *Inverse Document Frekuensi(Idf)*
4. Terakhir data akan di normalisasikan agar nilai tidak terpaut jauh dari data ke I sampai ke n

3.3 *Semi-Supervised K-Means*

Pada tahap ini dilakukan perhitungan untuk mengetahui hasil dari pengelompokan dokumen komentar di Tokopedia.



Gambar 3. Diagram Alir *Semi Supervised K-Means*

Dari Gambar 3. Seperti diatas maka dapat di artikan sebagai berikut :

1. Masukkan nilai dari hasil normalisasi pembobotan *TF-Idf* dan masukkan data label dari data komentar yang sudah berlabel.
2. Buat *variable labels* yang isinya adalah jumlah label yang bersifat unik dari data komentar Tokopedia.
3. Tentukan *centroid* dari data komentar yang sudah berlabel
4. Lakukan perhitungan jarak antara dokumen dengan *cluster* yang sudah ditentukan menggunakan rumus *CosineDistance*
5. Cari jarak yang nilai nya mendekati dengan letak *centroid* setelah itu simpan *index* dari

cluster tersebut ke dalam variable newCluster

6. Lakukan seleksi kondisi dimana apabila nilai *cluster* tidak terjadi perubahan maka program akan selesai jika masih ada perubahan kluster maka ulangi Langkah 4

4. PENGUJIAN DAN ANALISIS

Pada sub bab ini akan menjelaskan alur pengujian yang dilakukan pada penelitian ini. Pengujian ini untuk mengukur apakah metode-metode yang digunakan menghasilkan nilai yang diharapkan oleh penulis.pada pengujian kali ini akan dibagi menjadi 2 yaitu pengujian untuk mengukur nilai cluster yang terbentuk dan yang kedua adalah pengujian untuk hasil dari pengelompokan.

4.1 Pengujian nilai *Silhoutte*

Dalam pengujian yang akan dilakukan bertujuan untuk menghitung nilai dari *Silhoute Coefisien* sebagaimana yang telah dijelaskan pada Bab 2 sebelumnya. Pada percobaan kali ini akan memakai data sebanyak 300 data dengan memakai data latih sebanyak 50%,60%,70% dan 80% didapatkan hasil yang terdapat pada Tabel 2.

Tabel 2. Hasil Pengujian Menggunakan *Silhoutte*

Metode	Jumlah Data Latih	*Silhoutte*	*Structure*
Semi Supervised K-Means	50%	0.005499	*No Structure*
	60%	0.004869	*No Structure*
	70%	0.006088	*No Structure*
	80%	0.013944	*No Structure*
K-Means	50%	0.007354	*No Structure*
	60%	0.007198	*No Structure*
	70%	0.007276	*No Structure*
	80%	0.009825	*No Structure*

Pada perobaan pertama yang dilakukan menggunakan 50% data latih dari keseluruhan data dan 50% data sebagai data uji atau sebesar 150 untuk setiap data uji dan data latih didapatkan hasil bahwa metode *K-Means* memiliki nilai *Silhoutte* sebesar 0.006986 lebih besar dari pada metode *Semi Supervised K-Means* yang memiliki nilai *Silhoutte* sebesar 0.006269. Pada percobaan ke 2 untuk data latih sebesar 60% yang artinya jumlah data latih sebesar 180 data dan 120 data untuk data uji mendapatkan hasil *Silhoutte* pada masing-masing metode *Semi Supervised* dan Metode *K-Means* sebesar 0.005354 dan 0.006919. Pada percobaan ketiga untuk jumlah data latih sebesar 70% data yang artinya sebanyak 210 data menjadi data latih dan 90 data menjadi data uji didapatkan hasil *Silhoutte* untuk metode *K-Means* sebesar 0.006238 dan untuk metode *Semi Supervised* mendapatkan nilai *Silhoutte* sebesar 0.005887. Pada percobaan terakhir untuk jumlah data sebesar 80% data yang artinya 240 data menjadi data latih dan 60 data menjadi data uji didapatkan hasil *Silhoutte* sebesar 0.008756 untuk metode *K-Means* dan 0.013647 untuk metode *Semi Supervised K-Means*.

4.2 Analisis Perbandingan Hasil Pengelompokan Pada Metode *Semi Supervised K-Means* dan *K-Means.*

Pada pembahasan pertama untuk percobaan dengan 50% data sebgai data latih yang terdiri dari 150 data latih dan 150 data uji. Data yang diambil untuk perbandingan sebanyak 20 data dari data uji beserta label data yang benar seperti yang dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Hasil Pengelompokan

Jumlah Data Latih	Jumlah Benar		Persentase %	
	Semi Supervised K-Means	*K-Means Original*	*Semi Supervised K-Means*	*K-Means Original*
50%	95	70	63 %	46 %
60%	72	54	60 %	45 %
70%	46	31	51 %	34 %
80%	33	17	55 %	28 %

Dari hasil percobaan di atas dapat kita simpulkan bahwa dari hasil perbandingan yang telah dilakukan dapat disimpulkan untuk hasil pengelompokan dokumen ulasan menggunakan metode *Semi Supervised K-means* lebih unggul dari pada menggunakan metode *K-Means.*Hal

tersebut disebabkan oleh inisiasi *centroid* awal yang berbeda. Pada metode *Semi Supervised K-Means centroid* didapatkan dari mengambil nilai rata-rata setiap data berlabel pada data latih sehingga informasi term/kata yang didapatkan menjadi lebih banyak pada suatu label. Berbeda pada metode *K-Means* yang inisiasi *centroid* yang dilakukan berupa mengacak suatu nilai pada data latih yang mengakibatkan informasi term/kata menjadi terbatas dan kurang maksimal untuk mengelompokkan dokumen berdasarkan label yang ada.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari pengujian yang telah dilakukan pada penelitian Pengelompokan Ulasan Produk HP pada *Marketplace* Tokopedia Menggunakan Metode *Semi Supervised K-Mean* didapatkan Kesimpulan sebagai berikut :

1. Pada 4 percobaan yang telah dilakukan dengan membagi data latih menjadi 50%,60%,70%,80% dari keseluruhan data sebesar 300 data didapatkan evaluasi nilai *cluster* menggunakan *Silhouette* didapatkan hasil pada 3 percobaan awal nilai *Silhouette* yang dihasilkan lebih tinggi dengan menggunakan metode *k-means* sedangkan pada percobaan terakhir dengan 80% data nilai *Silhouette* yang dihasilkan pada metode *Semi Supervised K-Means* memiliki nilai yang lebih tinggi dari semua percobaan yang ada.
2. Pada penelitian ini nilai *Silhoutte* yang dihasilkan sangat kecil dikarenakan sebuah dimensi yang besar untuk mengukur keterdekatan dokumen sehingga perhitungan jarak kuran maksimal dan juga pada penelitian ini jumlah klaster yang sedikit juga menjadi penyebab bahwa nilai *silhouette* yang dihasilkan sangat kecil. Alasan mengapa pada metode *K-Means* memiliki nilai *Silhoutte* yang lebih baik adalah dikarenakan keterdekatan antar dokumen yang dihasilkan berdasarkan kata pembangunnya seperti {“barang”, ”pengiriman”, ”produk”} sehingga keanggotaan pada sebuah *cluster* terbentuk terhadap kemiripan kata-kata tersebut.
3. Untuk analisis perbandingan hasil *cluster* antara metode *Semi Supervised K-Means* dan *K-Means* untuk 4 percobaan yang dihasilkan, Hasil *cluster* yang dihasilkan pada metode *Semi Supervised K-Means* memiliki hasil yang lebih baik dan sesuai dengan label aslinya dikarenakan untuk inisiasi *centroid* awal yang dilakukan diambil dari nilai rata-rata dari data yang berlabel dibandingkan dengan metode *K-means* yang nilai *centroid* diacak sehingga hasil *cluster* yang dihasilkan tidak sesuai label aslinya.

5.2 Saran

Setelah dilakukan penelitian tentang pengelompokkan ulasan produk hp pada *marketplace* Tokopedia Adapun saran yang ditambahkan sebagai berikut :

1. Perlu ditambahkan sebuah metode *Preprocessing* data untuk mengatasi beberapa kata singkat atau kata yang tidak baku seperti {“brg”,”bgs”,”trmksh”}
2. Dalam tahap *Preprocessing* pada tahap *stopword removal* pada penelitian ini menggunakan *library* dari *sastrawi* dinilai kurang cukup untuk mengatasi kata seperti {“walaupun”,”kalaupun”,”semestinya”} untuk penelitian selanjutnya bisa digunakan sebuah *library* lain atau sebuah metode yang dapat mengatasi hal tersebut
3. Untuk penelitian selanjutnya dapat dilakukan uji sebuah *cluster* dengan jumlah *cluster* berjumlah 2 untuk melihat nilai *Silhoutte* yang dihasilkan apakah lebih baik daripada menggunakan *cluster* dengan jumlah 3

DAFTAR PUSTAKA

Assegaff, S. (2015). Pengaruh Trust (Kepercayaan) dan Online Shopping Experiences (Pengalaman Berbelanja Online) terhadap Perilaku Konsumen dalam Berbelanja Online: Prespektif Konsumen di Indonesia. *Jurnal Aplikasi Manajemen (JAM)*, *13*(Nomor 3), 463–473.

Gao, Y., Qi, H., Liu, D. Y., & Liu, H. (2008). Semi-supervised k-means clustering for multi-type relational data. *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, ICMLC*, *1*(July), 326–330. https://doi.org/10.1109/ICMLC.2008.4620425

Harahap, D. A. (2018). Perilaku Belanja Online Di Indonesia: Studi Kasus. *JRMSI - Jurnal Riset Manajemen Sains Indonesia*, *9*(2), 193–213. https://doi.org/10.21009/jrmsi.009.2.02

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529310. *Biometrics*, *33*(1), 159–174.

Li, Z., Jia, L., & Su, B. (2019). Improved K-Means Algorithm for Finding Public Opinion of Mount Emei Tourism. *Proceedings - 2019 15th International Conference on Computational Intelligence and Security, CIS 2019*, 192–196. https://doi.org/10.1109/CIS.2019.00048

Mahyoub, M., Hind, J., Woods, D., Wong, C., Hussain, A., & Aljumeily, D. (2019). Hierarchical text clustering and categorisation using a semi-supervised framework. *Proceedings - International Conference on Developments in ESystems Engineering, DeSE*, *October-20*, 153–159. https://doi.org/10.1109/DeSE.2019.00037

Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, *181*(1), 25–29. https://doi.org/10.5120/ijca2018917395

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saputra, T. I., & Arianty, R. (2019). Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat. *Jurnal Ilmiah Informatika Komputer*, *24*(3), 191–198. https://doi.org/10.35760/ik.2019.v24i3.2361

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

Wu, J., Du, L., & Dang, Y. (2018). Research on the Impact of Consumer Review Sentiments from Different Websites on Product Sales. *Proceedings - 2018 IEEE 18th International Conference on Software Quality, Reliability, and Security Companion, QRS-C 2018*, 332–338. https://doi.org/10.1109/QRS-C.2018.00065