ANALISIS SENTIMEN PADA BULETIN MENGGUNAKAN ALGORITME DBSCAN

Dwi Vina Wijaya[1*)], Yogi Firman Alfiansyah[1)], Anton Romadoni Junior[1)], Anis Fitri Nur Masruriyah[1)], Jamaludin Indra[1)], Hanny Hikmayanti[1)], Amril Mutoi Siregar[1)]

Teknik Informatika, Universitas Buana Perjuangan, Karawang
*Email Korespondensi: If17.dwiwijaya@mhs.ubpkarawang.ac.id

ABSTRAK

Asosiasi Perguruan Tinggi Informatika dan Ilmu Komputer (APTIKOM) mengembangkan media yang memuat informasi dan teknologi bentuk media online yaitu Buletin. Informasi yang disampaikan pada Buletin membahas mengenai big data, kemajuan teknologi dan lain – lain. Kalimat yang terkandung dalam Buletin dapat berupa kalimat positif, negatif maupun netral. Penggunaan kata dalam menyusun kalimat dapat memengaruhi informasi yang disampaikan. Oleh karena itu, penyusunan kalimat perlu diperhatikan agar dapat meminimalkan kesalahan maksud dan tujuan. Penyusunan kalimat dapat diawali dengan pemilihan kata, pengelompokkan kata, dan melakukan klasifikasi sentimen. Proses pemeriksaan dokumen dapat dilakukan dengan algoritme DBSCAN. Algoritme DBSCAN dapat melakukan clustering dalam menentukan noise yang terdapat di dalam dokumen. Penelitian magazine bertujuan melakukan pemeriksaan kata negatif, positif dan netral. Selain itu, bertujuan untuk melakukan pencarian intisari yang terdapat dalam dokumen. Tahapan diawali dengan proses TF IDF untuk klasifikasi dan DBSCAN untuk clustering. Selanjutnya, hasil yang diperoleh akan dievaluasi dengan Sum of Square Error (SSE) dan pemeriksaan ketepatan cluster meggunakan Silhouette. Hasil evaluasi algoritma menunjukkan perbandingan nilai masing – masing cluster. Lalu, hasil evaluasi akan diperiksa dengan silhouette yang menunjukkan ketepatan cluster.

Kata kunci: Buletin, DBSCAN, Literasi, Text Mining

ABSTRACT

The Association of Informatics and Computer Science Universities (APTIKOM) develops media that contains information and technology in online media, namely Bulletins. The information presented in the Bulletin discusses big data, technological advances, and others. Sentences contained in the Bulletin can be positive, negative or neutral sentences. The use of words in constructing sentences can affect the information conveyed. Therefore, the arrangement of sentences needs to be considered to minimize errors of intent and purpose. Sentence structure can be started by selecting words, grouping words, and classifying sentiments. The document checking process was conducted by using the DBSCAN algorithm. The DBSCAN algorithm can perform clustering in determining the noise contained in the document. Magazine research aims to examine negative, positive and neutral words. In addition, it aims at finding the essence contained in the document. The stage begins with the TF IDF process for classification and DBSCAN for clustering. Furthermore, the results obtained will be evaluated with the Sum of Square Error (SSE) and checked the accuracy of the cluster using Silhouette. The results of the evaluation of the algorithm show a comparison of the values of each cluster. Then, the evaluation results will be checked with a silhouette that shows the accuracy of the cluster.

Keywords: *Bulletin, DBSCAN, Literacy, Text Mining*

PENDAHULUAN

Literasi menjadi hal pokok yang perlu dimiliki setiap orang, khususnya kemampuan menulis, membaca, dan berhitung. Literasi memiliki tujuan untuk menyampaikan

informasi, memecahkan masalah, menulis laporan, melakukan penelitian dan lain – lain. Indonesia memiliki 750 juta remaja dan orang dewasa belum bisa membaca serta menulis, dengan 250 juta anak gagal memperoleh keterampilan keaksaraan [1]. Keterampilan keaksaraan Indonesia memperoleh peringkat 60 dari 61 negara dari minat baca yang sangat rendah [2]. Oleh karena itu, untuk meningkatkan minat baca, Asosiasi Perguruan Tinggi Informatika dan Ilmu Komputer (APTIKOM) membuat media literasi daring yang dapat diakses oleh pengguna internet. Media yang dikembangkan oleh APTIKOM yaitu *magazine* Buletin yang memuat informasi seputar teknologi dan visualisasi data. *Magazine* Buletin merupakan salah satu contoh literasi *online* selain bentuk fisik buku pengetahuan, buku cerita dan lain lain. Informasi yang disampaikan *magazine* Buletin dapat berupa pengetahuan *software*, *hardware*, teknologi, data, dan lain – lain. Terdapat hal – hal yang perlu diperhatikan dalam menyampaikan informasi seperti penggunaan kata yang baik dan benar, serta visualisasi gambar. Informasi yang disampaikan oleh *magazine* Buletin dapat mengandung tanggapan yang berbeda – beda oleh pembaca. Sehingga, menimbulkan berbagai macam sentimen yang bernilai positif, negatif, dan netral. Oleh karena itu, dalam menyampaikan informasi *magazine* Buletin perlu melakukan pemeriksaan kata untuk meminimalkan kesalahan penulisan dan penggunaan kata.Hal ini berdampak pada tanggapan yang disampaikan oleh pembaca mengenai *magazine* Buletin.

Meminimalkan kesalahan dapat diselesaikan berdasarkan kandidat data yang diberikan seperti menghapus satu huruf, menambah satu huruf dan lain – lain [3]. Salah satu hal untuk menghindari berbagai macam reaksi pembaca dapat melakukan analisis sentimen positif, negatif, dan netral. Hal – hal yang dilakukan dimulai dari pemeriksaan kata, mengelompokkan kata dokumen dan membagi kata sesuai bobot nilai. Bobot nilai yang diperoleh merupakan hasil dari perhitungan dan pengelompokkan kata sering muncul dalam dokumen. Kata yang memiliki bobot nilai tertinggi merupakan intisari dari dokumen. Proses pengelompokkan kata akan menghasilkan kata sering muncul, intisari dokumen, dan sentimen dari kata yang terkandung dalam dokumen. Proses analisis sentimen (AS) teks dalam dokumen tidak ditentukan berdasarakan spesifik label dan entitas target dengan kata lain [4]. Analisis sentimen akan menghasilkan nilai sentimen positif, negatif, dan netral dari masing – masing kata dalam dokumen serta akan menghasilkan intisari dokumen. Proses penambangan teks atau intisari dokumen dapat menggunakan bahasa pemrograman R dengan berbagai macam *tools* seperti *tools statistic* dari *linear* memodel non *linear*, analisis *time-series*, klasifikasi, *clustering* dan lain – lain [5]. Bahasa pemrograman R dapat dilakukan menggunakan *software* RStudio dalam menyelesaikan permasalahan seperti analisis sentimen, mencari intisari dalam dokumen dan lain – lain. Paket-R adalah kumpulan fungsi R yang merupakan hasil kompilasi kode pada data sampel [6]. Rstudio memiliki fungsi yang dapat melakukan ekstraksi, pembagian, perhitungan, visualisasi data dan lain – lain. Hasil yang diperoleh dengan RStudio akan menampilkan visualisasi data dalam bentuk awan kata atau *wordclouds*. *Wordclouds. Wordclouds* atau awan kata adalah sejenis daftar berbobot untuk memvisualisasikan bahasa atau data teks [7].

METODE PENELITIAN

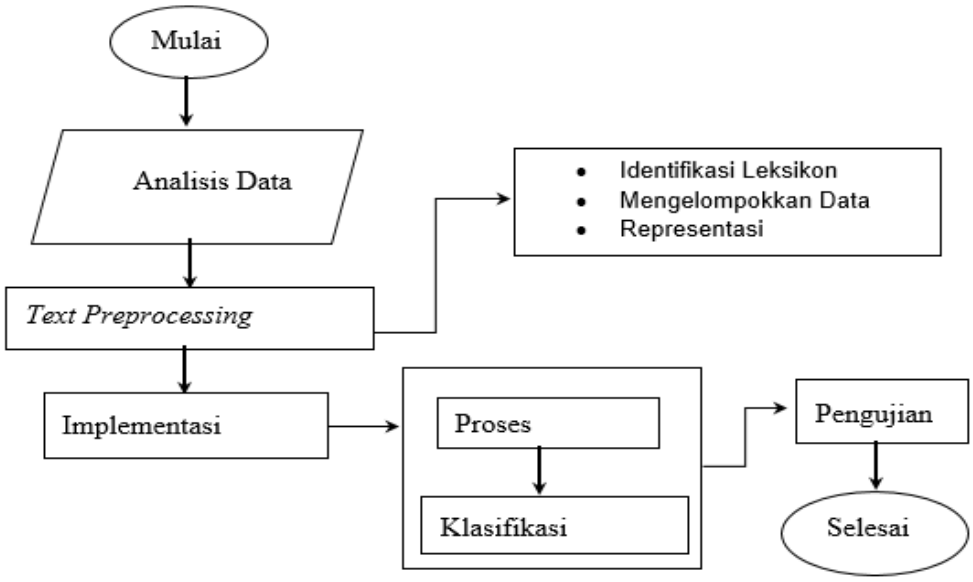
A. Bahan dan Peralatan

Data yang digunakan dalam melakukan bahan penelitian yaitu data *magazine* Buletin 2019/2020 yang memuat informasi seputar teknologi yang dikembangkan oleh APTIKOM. Selain itu, terdapat bahan yang dijadikan sebagai referensi penelitian analisis sentimen yaitu buku dan studi literatur. Adapun peralatan yang digunakan dalam analisis sentimen diantaranya:

1. Perangkat Keras
   - laptop dengan spesifikasi AMD *Ryzen* 5 2500U *with Radeon Vega Mobile*, 1 *Terabyte*, 8 GB RAM, *mouse*, *flashdisk*.
2. Perangkat Lunak
   - Rstudio
   - *Microsoft Word*
   - *Microsoft Excel*

B. Prosedur Penelitian

Prosedur Penelitian diawali dengan melakukan analisis data *magazine* Buletin untuk memperoleh data dan mendapatkan informasi. Setelah itu, data *magazine* akan dilakukan proses *text preprocessing* untuk melakukan pengelompokkan kata sampai dengan tahap akhir yaitu pengujian. Prosedur penelitian ditunjukkan pada gambar 1.



Gambar 1. Prosedur penelitian

Penelitian sentimen dimulai dengan melakukan analisis data dilakukan untuk mengetahui permasalahan yang terdapat pada *magazine* Buletin dengan mengambil beberapa sampel kalimat yang tertera pada *magazine* Buletin. Lalu, dilakkukan tahapan *text preprocessing* untuk mengetahui aktualitas, kelengkapan data dengan melakukan identifikasi leksikon, mengelompokkan data, dan melakukan representasi data. Kemudian, melakukan implementasi algoritme untuk mengetahui proses perbandingan nilai, dan pengelompokkan sentimen. Setelah itu, dilakukan pengujian untuk mengetahui tingkat akurasi algoritme menggunakan metode *Sum of Square Error* (SSE) untuk menguji performa algoritme DBSCAN. Hal ini ini dilakukan untuk mengetahui tanggapan masyarakat terhadap *magazine* buletin melalui penggunaan kata.

C. Akuisisi Data

Data didapatkan dari *magazine* Buletin 2019/2020 secara *online* yang akan dijadikan sebagai data awal rujukan dalam melakukan proses analisis sentimen yang memuat 191 data. Data *magazine* Buletin dilakukan tahapan menggunakan sistem untuk mencari kata sering muncul, sentimen positif, negatif dan netral, serta visualisasi data. Berikut data awal *magazine* Buletin yang telah dimasukkan ke dalam sistem ditunjukkan pada tabel 1.

ISSN Cetak : 2622-1276
ISSN Online : 2622-1284

Tabel 1. Data awal *magazine* Buletin

No	Kalimat
1	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan?"
2	"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan?"
3	"Big data seringkali digunakan untuk tujuan memprediksi sesuatu dalam berbagai bidang ilmu namun seringkali hasilnya jauh dari ekspektasi."
4	"Bisakah Big Data meningkatkan akurasinya?"
5	"Untuk membantu mempercepat pembangunan dan melakukan tindakan preventif terhadap bencana yang akan datang, Big data telah diterapkan di berbagai sektor."
.	-
.	-
.	-
187	"10",
188	"ABOUT",
189	"Copyright © 2020",
190	"Buletin Aptikom",
191	"All rights are reserved."

D. Text Pre processing

Text pre processing merupakan tahapan yang dilakukan untuk analisis data dengan melakukan penyeleksian data dan mengubah data tidak terstruktur menjadi data terstruktur. Tahap pra pemrosesan akan menghasilkan pengelompokkan kata, penyeleksian, menghitung bobot nilai setiap kata yang sering muncul dan mencari intisari dari dokumen. Tahapan yang dilakukan diantaranya**:**

1) Identifikasi Leksikon (kata)
   Identifikasi leksikon atau kata dilakukan untuk melakukan penyaringan kata dari dokumen untuk menetapkan parameter dari setiap pengelompokkan kata.
2) Mengelompokkan Data
   Mengelompokkan kata dilakukan untuk mengetahui dan membagi dari masing – masing kata yang akan didistribusikan sesuai dengan parameter sentimen
3) Representasi
   Representasi dilakukan untuk melakukan visualisai data yang telah dikelompokkan dan pembagian.

Hasil yang diperoleh akan dilakukan proses implementasi algoritme untuk mengetahui data pencilan yang terdapat pada dokumen.

E. Algoritme DBSCAN

Implementasi algoritme menggunakan data *preprocessing* untuk melakukan tahap *clustering* dengan algoritme DBSCAN untuk mengetahui pencilan dari *magazine* Buletin. Proses *clustering* diawali dengan menentukan minimum poin dan jarak yang akan menentukan sebaran data dan untuk mengatahui pencilan atau *outliers*. Algoritme DBSCAN memiliki tiga jenis poin diantaranya poin inti (*core*), poin tepi (*border*), dan poin pencilan (*outlier*). Poin yang memenuhi jarak radius ε (epsilon) dan minimum poin disebut dengan poin inti. Apabila poin tetangga (*neighbor*) memenuhi jarak dan minimum poin dekat dengan *cluster* poin atau poin inti maka disebut dengan poin tepi (*border*). Namun, jika terdapat poin tunggal yang tidak memenuhi jarak epsilon dan minimum poin maka disebut poin pencilan (*outliers*). Algoritme DBSCAN ditentukan oleh pemilihan parameter Eps (ε) dan MinObj yang optimal dipilih berdasarkan k-dist *graph* [7]. Adapun beberapa tahapan yang dilakukan diantaranya:

1) Menentukan jarak epsilon dan minimum poin (minPts) pada sebaran data

2) Menentukan poin inti secara acak pada sebaran data yang akan dijadikan poin acuan dalam menentukan *cluster*
3) Menghitung jarak *epsilon* poin inti dengan poin tetangga menggunakan *Euclidean Distance*
4) Menghitung jumlah poin yang memenuhi jarak dan minPts pada poin acuan
5) Poin yang memenuhi jarak dan melebihi minPts maka poin acuan ditandai dengan poin inti dan membentuk suatu *cluster*
6) Apabila terdapat poin yang kurang memenuhi minimum poin dan jarak radius pada poin acuan maka disebut dengan poin tepi (*border*)
7) Jika pada sebaran data terdapat poin yang tidak memenuhi minPts dan radius serta poin terletak jauh dengan sebaran data atau disebut poin tunggal maka poin tersebut ditandai sebagai poin pencilan (*outliers*).
8) Ulangi tahapan yang dilakukan secara *iterative* pada sebaran data lainnya.

Jarak *epsilon* pada poin satu dengan poin tetangga (ε - *neighborhood*) dapat dihitung menggunakan rumus *Euclidean Distance* dengan persamaan 1 sebagai berikut [8]:

$$d_{ij} = \sqrt{\sum_{a=1}^{p} \left(x_{ia} - x_{ja}\right)^2}, i = 1, \dots, n; j = 1, \dots, n \qquad 1.$$

Keterangan:
x_{ia} = variabel ke a dari objek i $(i = 1, \dots, n; a = 1, \dots, p)$
d_{ij} = nilai *Euclidean Distance*
N = banyaknya fitur

Menentukan nilai minPts dan jarak radius ε sebagai parameter dari sebaran data maka k – *dist graph* dapat digunakan untuk mendapat nilai minPts dan radius ε *epsilon*. Sebelum melakukan tahap *clustering*, diperlukan proses pengelompokkan kata untuk mengetahui pola sering muncul, menganalis sentimen, dan intisari menggunakan formula TF IDF. Berikut tahapan yang dilakukan:

1) Proses
Proses implementasi dimulai dengan melakukan pengelompokkan kata dengan menerapkan formula TF – IDF untuk mengetahui perbandingan bobot nilai dari masing – masing kata. Selain itu, proses TF IDF akan menghasilkan pola sering muncul yang menjadi intisari dari dokumen. Berikut persamaan 2, 3, dan 4 untuk membentuk formula TF IDF (Isnarwaty & Irhamah, 2019).

$$TF_j = \left(\frac{i}{DF_j}\right) \qquad (1)$$

$$IDF_j = log\ log\ \left(\frac{N}{DF_j}\right) \qquad (2)$$

$$W_{i,j} = TF_{i,j} \times IDF_j, \qquad (3)$$

Keterangan:
i = *term*
j = dokumen
$W_{i,j}$ = bobot dari kata ke *j* pada kata *i*
IDF*j* = *inverse document frequency* pada kata ke *j*
TF*i,j* = jumlah kemunculan kata ke *j* pada kata ke *i*
DF*j* = banyaknya kalimat yang mengandung kata *j*
N = jumlah keseluruhan kata

2) Klasifikasi
   Klasifikasi dilakukan untuk mengetahui setiap kata yang telah melakukan tahap pengelompokkan sesuai dengan bobot nilai yang diperoleh masing – masing kata.

3) Evaluasi
   Tahap evaluasi dilakukan untuk mengetahui kesesuaian pengelompokkan dari masing – masing kata terhadap kelas sentimen yang didistribusikan.

Tahap selanjutnya yang dilakukan yaitu melakukan *clustering* untuk mengetahui data pencilan yang ada pada dokumen.

HASIL DAN PEMBAHASAN

Penelitian sentimen menghasilkan data *text pre processing* dan hasil *clustering* menggunakan algoritme DBSCAN yang memuat hasil sentimen positif, negatif , netral dan intisari dokumen, serta data DBSCAN *clustering*. Berikut data yang telah melakukan beberapa tahapan:

A. Data hasil Text Preprocessing

Data *preprocessing* atau pra pemrosesan data didapatkan dari proses implementasi *magazine* Buletin ke dalam sistem sebanyak 191 data. Tahapan *text preprocessing* diantaranya *To lower* , *removeNumber*, *removePunctuation*, *stripWhitespace*, *removeWords, stopwords*, *stemDocument*. Hasil implementasi *text preprocessing* menghasilkan data pada tabel 2.

Tabel 2 Tabel hasil *text preprocessing*

No	Proses	Sebelum	Sesudah
1	*To lower*	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?",	apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan
		"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?",	lalu pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan
2	*removeNumber*	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?",	apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan ?
		"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?",	lalu, pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan ?
3	*removePunctuation*	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?",	apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan
		"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?",	lalu pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan
4	*stripWhitespace*	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?",	apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan
		"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?",	lalu pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan
5	*removeWords, stopwords*	"Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?",	apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan

6	*stemDocument*.	"Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?", "Apa yang dimaksud dengan Big Data dan dalam bidang apa sajakah Big Data itu digunakan ?", "Lalu, pengaruh apa yang di capai oleh bantuan Big Data dalam 5 sampai 10 tahun ke depan ?",	lalu pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan apa yang dimaksud dengan big data dan dalam bidang apa sajakah big data itu digunakan lalu pengaruh apa yang di capai oleh bantuan big data dalam sampai tahun ke depan

B. Implementasi Algoritme

Data yang digunakan untuk implementasi algoritme yaitu data *magazine* Buletin 2019/2020 dalam melakukan analisis sentimen. *Magazine* Buletin membahas mengenai kegunaan big data dalam revolusi 4.0 yang dapat diterapkan dalam berbagai sektor. Sampel data yang digunakan untuk *clustering* algoritme DBSCAN sebanyak 191 data dengan menetapkan jarak dan minimum poin untuk mengetahui sebaran data. Tahapan diawali dengan *tidy text* dengan sepuluh sampel data pada gambar 2, 3 dan 4.

	Nomor	Kalimat
1	1	apa yang dimaksud dengan big data dan dalam bidang apa ...
2	2	lalu pengaruh apa yang di capai oleh bantuan big data dala...
3	3	big data seringkali digunakan untuk tujuan memprediksi ses...
4	4	bisakah big data meningkatkan akurasinya
5	5	untuk membantu mempercepat pembangunan dan melaku...
6	6	intermezzo
7	7	pada era revolusi industri yang semakin maju ini banyak aka...
8	8	kegunaan big data
9	9	rumah sakit dapat merekam catatan medi pasien sehingga ...
10	10	physic medic sain

Gambar 2. Data awal Buletin

	nomor	baris	word
1	1	1	apa
2	2	1	yang
3	3	1	dimaksud
4	4	1	dengan
5	5	1	data
6	6	1	dan
7	7	1	dalam
8	8	1	bidang
9	9	1	apa
10	10	1	sajakah

Gambar 3. Data pengelompokka *tidy text*

Data awal *magazine* Buletin kemudian dilakukan pengelompokkan, pencarian kata sering muncul menggunakan TF IDF diawali dengan proses *tidy text* ditunjukkan pada gambar

	word	n
1	data	156
2	yang	79
3	dan	50
4	gambar	31
5	dari	30
6	dengan	26
7	pada	25
8	adalah	24
9	atau	22
10	dalam	21

Gambar 4 Frekuensi data *tidy text*

Setelah melakukan pengelompokkan, tahap selanjutnya yaitu analisis sentimen dari setiap masing – masing kata dengan memperoleh hasil sentimen ditunjukkan pada gambar 6.

```
negative  neutral positive
       8     1917       25
```

Gambar 5. Jumlah sentimen positif, negatif, dan netral *magazine* Buletin

	word	emotion	polarity
1	apa	anger	neutral
2	yang	anger	neutral
3	dimaksud	anger	neutral
4	dengan	anger	neutral
5	data	anger	neutral

Gambar 6. Hasil sentimen sentimen positif, negatif, dan netral

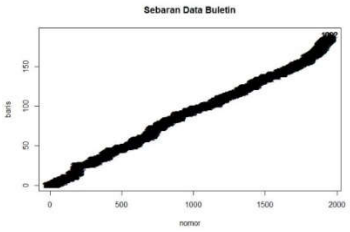
ISSN Cetak : 2622-1276
ISSN Online : 2622-1284

Adapun pengelompokkan kata memperoleh hasil TF IDF dengan masing – masing kata yang telah dilakukan pengelompokkan.

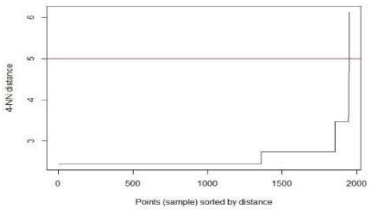
	word	Nomor	wordTotal	n	tf	idf	tf_idf
1	data	141	156	225	0.080271138	1.4663371	0.11770455
2	akan	11	16	124	0.318766067	1.3121864	0.41828049
3	banyak	103	15	120	0.324324324	0.9555114	0.30989360
4	dan	103	50	120	0.106288751	0.9555114	0.10156012
5	data	103	156	120	0.042811274	0.9555114	0.04090666
6	data	49	156	102	0.036389583	0.8602013	0.03130236
7	data	98	156	102	0.036389583	0.9555114	0.03477066
8	data	91	156	96	0.034249019	1.3121864	0.04494110
9	dan	73	50	90	0.079716563	1.8718022	0.14921364
10	data	50	156	90	0.032108455	1.1786550	0.03784479

Gambar 7. Hasil TF IDF

Tahapan selanjutnya, yaitu proses *clustering* untuk mengetahui poin pencilan dari data *magazine* buletin yang ditunjukkan pada gambar 9.

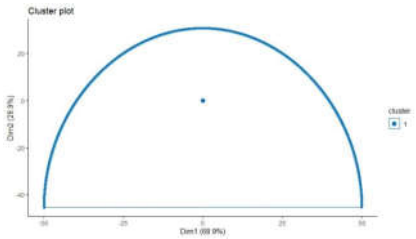


Gambar 8. Sebaran data buletin



Gambar 9. K – Dist *graph* sebaran data berdasarkan jarak

K – dist *graph* menunjukkan *distance* dengan parameter jarak epsilon 5 dan minimum poin 4, proses *clustering* menghasilkan satu *cluster* dengan 0 *noise* poin ditunjukkan pada gambar 11.



Gambar 10. *Cluster* data Buletin menggunakan DBSCAN

C. Evaluasi

Berdasarkan tahapan yang telah dilakukan dengan algoritme DBSCAN, maka data hasil *cluster* DBSCAN akan dievaluasi menggunakan *Sum of Square Error* (SSE). Sampel data yang digunakan yaitu sepuluh sampel dari 1950 data dengan tipe data data numeric *magazine* buletin. Berikut Hasil evaluasi menggunakan SSE yang ditunjukkan pada gambar 12.

	nomor	baris
1	-1.730719	-1.778910
2	-1.728943	-1.778910
3	-1.727167	-1.778910
4	-1.725391	-1.778910
5	-1.723615	-1.778910
6	-1.721839	-1.778910
7	-1.720063	-1.778910
8	-1.718287	-1.778910
9	-1.716511	-1.778910
10	-1.714735	-1.778910

Gambar 11 Data Numerik Buletin

Data yang diperoleh dengan evaluasi menggunakan SSE menghasilkan nilai tengah dan nilai evaluasi 88.6 % dari proses pengelompokkan kata. Berikut hasil evaluasi nilai tengah dan hasil evaluasi *cluster* SSE menggunakan algoritme DBSCAN, K Means pada gambar 12 dan 14. Serta, hasil evaluasi menggunkan Confusion Matrix dengan algoritme K Nearest Neighbors pada gambar 15.

```
Within cluster sum of squares by cluster:
[1] 184.7372 129.3869 128.8614
 (between_SS / total_SS =  88.6 %)
```

Gambar 12 Nilai evaluasi SSE dengan algoritme DBSCAN

```
$centers
        nomor      baris
1  0.08169632  0.1200608
2 -1.11444442 -1.1296227
3  1.19614074  1.1721528

$totss
[1] 3898

$withinss
[1] 128.8614 184.7372 129.3869

$tot.withinss
[1] 442.9855

$betweenss
[1] 3455.014
```

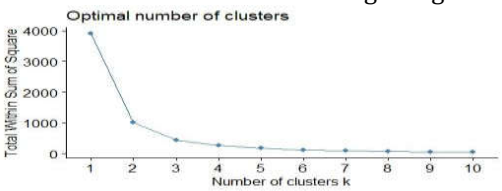
Gambar 13 Nilai tengah evaluasi SSE

```
Within cluster sum of squares by cluster:
[1] 15448.80 16259.59
 (between_SS / total_SS =  75.0 %)
```

Gambar 14 Nilai evaluasi SSE dengan algoritme K Means

```
Confusion Matrix and Statistics

          Reference
Prediction negative neutral positive
  negative        0       0        0
  neutral         0       0        1
  positive        1       2       25

Overall Statistics

               Accuracy : 0.8621
                 95% CI : (0.6834, 0.9611)
    No Information Rate : 0.8966
    P-Value [Acc > NIR] : 0.8249

                  Kappa : -0.045

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: negative Class: neutral Class: positive
Sensitivity                  0.00000        0.00000          0.9615
Specificity                  1.00000        0.96296          0.0000
Pos Pred Value                   NaN        0.00000          0.8929
Neg Pred Value               0.96552        0.92857          0.0000
Prevalence                   0.03448        0.06897          0.8966
Detection Rate               0.00000        0.00000          0.8621
Detection Prevalence         0.00000        0.03448          0.9655
Balanced Accuracy            0.50000        0.48148          0.4808
```

Gambar 15 Nilai evaluasi Confusion Matrix dengan algoritme K Nearest Neighbors



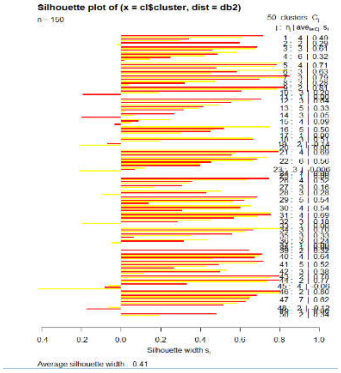
Gambar 16 Sebaran data total dalam *cluster* SSE

Data yang diperoleh setelah melakukan evaluasi menghasilkan data sebaran lebih dari 3000 data. Selanjutnya, data yang telah dievaluasi akan dilakukan proses pemeriksaan *cluster* menggunakan *Silhouette* dengan 50 sampel data iris SSE.

```
within cluster sum of squares by cluster:
 [1] 0.00000000 0.21250000 0.09000000 0.14000000
[12] 0.00000000 0.06000000 0.04500000 0.02500000
[23] 0.03333333 0.07000000 0.11250000 0.05500000
[34] 0.33500000 0.29500000 0.10000000 0.42000000
[45] 0.15250000 0.04666667 0.00000000 0.19600000
 (between_SS / total_SS =  99.2 %)
```

Gambar 17 Total pemeriksaan *Silhouette*

Hasil yang didapatkan setelah dievaluasi menggunakan *silhouette* dengan *cluster* SSE memperoleh nilai sebesar 99.2 % dari pemeriksaan *cluster*. Nilai yang diperoleh menunjukkan akurasi *cluster* sesuai dengan hasil klasifikasi dan pengelompokkan kata pada *magazine* Buletin dengan *plot* yang ditunjukkan pada gambar 17.



Gambar 18 *Silhouette* plot *magazine* Buletin

ISSN Cetak : 2622-1276
ISSN Online : 2622-1284

KESIMPULAN

Berdasarkan penelitian yang dilakukan pada analisis sentimen pada buletin menghasilkan pengelompokkam kata sentimen positif, negatif, dan netral serta intisari dokumen. Hasil yang diperoleh menghasilkan sentimen 25 positif, 8 negatif, dan 1917 netral dengan intisari dari dokumen yaitu kata "data" dengan memperoleh nilai tertinggi kata sering muncul. Hasil evaluasi *Sum of Square Error* (SSE) *magazine* Buletin menunjukkan nilai akurasi 88.6% pengelompokkan kata menggunakan algoritme DBSCAN. Nilai akurasi nilai SSE bernilai 75.0% menggunakan K Means dan Hasil evaluasi confusion matrix sebesar 86.2% menggunakan K Nearest Neighbors. Selanjutnya, pemeriksaan *cluster* menghasilkan nilai 99.2 % dari hasil pemeriksaan *cluster*. Nilai akurasi dan pemeriksaan *cluster* menunjukkan ketepatan pengelompokkan kata dan proses *cluster* dari *magazine* Buletin. Hal ini menunjukkan bahwa tanggapan dapat dimulai dengan memerhatikan penggunaan kata. Adanya analisis meminimalkan tanggapan yang diberikan terhadap informasi yang disampaikan melalui penggunaan kata.

REFERENSI

[1] Evita Devega. (2017, Oktober) Kominfo. [Online]. Https://Www.Kominfo.Go.Id/Content/Detail/10862/Teknologi-Masyarakat-Indonesia-Malas-Baca-Tapi-Cerewet-Di-Medsos/0/Sorotan_Media

[2] H Richard Gail, Sidney L Hantler, And Dkk, "Method And Apparatus For Automatic Detection Of Spelling Errors In One Or More Documents," 2016.

[3] Nurulhuda Zainuddin, Ali Selamat, And Roliana Ibrahim, "Hybrid Sentiment Classification On Twitter Aspect-Based Sentiment Analysis," Crossmark, 2018.

[4] M Reza Faisal, Seri Belajar Pemrograman: Pengenalan Bahasa Pemrograman R. Banjarmasin, 2016.

[5] Vaddadi Vasudha Rani And K Sandhya Rani, "Twitter Streaming And Analysis Through R," Indian Journal Of Science And Technology, 2016.

[6] Yuping Jin, "Development Of Word Cloud Generator Software Based On Python," Elsevier, 2017.

[7] Windy Rohalidyawati, Rita Rahmawati, And Mustafid , "Segmentasi Pelanggan E-Moneydengan Menggunakan Algoritma Dbscan (Density Based Spatial Clustering Applications With Noise)Di Provinsi Dki Jakarta," Jurnal Gaussian, 2020.

[8] Muhammad Tanzil Furqon And Lailil Muflikhah, "Clustering The Potential Risk Of Tsunami Using Density-Based Spatial Clustering Of Application With Noise (Dbscan)," Journal Of Environmental Engineering & Sustainable Technology, P. 8, 2016.

[9] (2019) Unesco. [Online]. Https://En.Unesco.Org/Themes/Literacy

[10] H Richard Gail, Sidney L Hantler, And Dkk, "Method And Apparatus For Automatic".

[11] Richard Gail, Sidney L Hantler, And Meir M Laker, "Method And Apparatus For Automatic Detection Of Spelling Errors In One Or More Documents ," United States Patent, 2016.

[12] Feri Sulianta, Literasi Digital, Riset, Perkembangannya & Perspektif Social Studies. Bandung, 2020.

[13] Devi Putri Isnarwaty And Irhamah, "Text Clustering Pada Akun Twitter Layanan Ekspedisi Jne, J&T, Dan Pos Indonesia Menggunakan Metode Density-Based Spatial Clustering Of Applications With Noise (Dbscan) Dan K-Means," Jurnal Sains Dan Seni Its, 2019.