

Aplikasi Text Mining untuk Klasterisasi Aduan Masyarakat Kota Semarang Menggunakan Algoritma K-means

Dita Afida¹, Erika Devi Udayanti², Etika Kartikadharma³

¹Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang, e-mail: ditaafida@gmail.com

²Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang, e-mail: erikadevi@dsn.dinus.ac.id

³Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang, e-mail: etika.kartikadharma@dsn.dinus.ac.id

ARTICLE INFO

Article history:

Received 11 June 2020

Received in revised form 06 July 2020

Accepted 22 January 2021

Available online 31 January 2021

ABSTRACT

Social media is a service that is very supportive for government activities, especially in providing openness and community-based government. One form of its implementation is the Semarang City government through the Pusat Pengelolaan Pengaduan Masyarakat (P3M), whose task is to manage community complaints that enter one of the communication channels namely social media twitter. The number of public complaints that enter every day is very varied. This is certainly quite difficult for managers in categorizing complaints reports according to the relevant government organizations. This paper focuses on the problem of how to conduct clustering of community complaints. The data source comes from Twitter using the keyword "Laporhendi". Text document data from community complaint tweets was analyzed by text mining methods. This method begins with preprocessing data such as case folding, tokenizing, stemming, stopword removal and tf-idf measurement. Clustering algorithm that will be used in dividing the complaint cluster is k-means algorithm. The evaluation of cluster results is done by using purity to determine the accuracy of the clustering.

Keywords: community complaint, text mining, kmeans algorithm, purity

1. Pendahuluan

Masyarakat saat ini sangat akrab dengan keberadaan teknologi informasi dan komunikasi dalam aktivitas sehari-hari, baik aktivitas berbelanja, aktivitas sosial dan termasuk aktivitas di lingkungan pemerintahan. Adaptasi teknologi di lingkungan pemerintah dilakukan untuk memenuhi kebutuhan transparansi kerja pemerintah kepada masyarakat yang tentunya tujuan utamanya adalah dalam upaya peningkatan layanan publik yang dilakukan melalui penggunaan

Received June 11, 2020; Revised July 06, 2020; Accepted January 22, 2021

teknologi informasi dan komunikasi. Saat ini pemerintah pusat hingga daerah membuka kanal komunikasi elektronik secara luas bagi masyarakat untuk dapat berbagi informasi maupun menghimpun berbagai saran maupun aduan terkait layanan pemerintah. Kanal komunikasi antara pemerintah dan masyarakat ini disediakan dalam berbagai bentuk mulai dari web yang dapat diakses masyarakat selama 24 jam, aplikasi berbasis *mobile*, layanan berbasis *short message* hingga media sosial seperti twitter, facebook hingga instagram. Masyarakat menjadi semakin mudah memperoleh informasi dan membagikan informasi kepada masyarakat lainnya melalui internet. Salah satunya adalah yang berkaitan dengan aduan masyarakat berkaitan dengan layanan pemerintahan.

Dalam penelitian ini mengambil studi kasus pada Pusat Pengelolaan Pengaduan Masyarakat (P3M) khususnya di kota Semarang. Dimana kantor P3M ini bertugas untuk melakukan pengelolaan aduan masyarakat yang masuk. Banyaknya aduan masyarakat yang masuk tiap harinya sangatlah bervariasi. Hal ini tentunya cukup menyulitkan pengelola dalam mengkategorikan laporan pengaduan masyarakat. Dalam proses pengkategorisasian aduan, bidang pelaporan harus membaca satu persatu dan tentu memahami isi dari aduan masyarakat. Baru kemudian petugas menentukan aduan tersebut ditujukan ke unit pemerintah terkait untuk tindak lanjut. Mempertimbangkan jumlah aduan yang masuk setiap harinya, tentu proses pengkategorisasian tersebut memakan banyak waktu. Hal ini membuat semakin lama pula aduan tersampaikan dan ditindak lanjuti oleh Organisasi Pemerintah Daerah (OPD) yang terkait. Dari hasil wawancara yang dilakukan dengan pihak pengelola aduan masyarakat, ada sejumlah 11 Organisasi Pemerintah Daerah yang berkaitan dengan laporhendi, diantaranya yaitu OPD PLN, Dinas Perhubungan (Dishub), Dinas Perumahan dan Kawasan Permukiman (Disperkim), Pekerjaan Umum (PU), Dinas Kesehatan (Dinkes), OPD PDAM, OPD Penanggulangan Bencana Daerah (BPBD), OPD Dinas Kependudukan Pencatatan Sipil (Dukcapil), Rumah Sakit Umum Daerah (RSUD), Satuan Polisi Pamong Praja (Satpol PP) dan terakhir OPD Dinas Lingkungan Hidup. Pengelola harus melakukan pengkategorisasian setiap aduan masyarakat yang masuk ke sejumlah OPD tersebut.

Penelitian ini fokus pada permasalahan bagaimana melakukan pengelompokan (*clustering*) aduan masyarakat yang disampaikan melalui twitter. *Clustering* sendiri merupakan salah satu metode dalam data mining yang melakukan pemetaan terhadap sejumlah besar data menjadi klaster atau kelompok yang berbeda dan memasukkannya ke dalam klaster yang sama satu dengan yang lain dan yang berbeda dari klaster lain. Sumber data berasal dari data aduan masyarakat yang diambil dari media sosial twitter yang berupa dokumen text dengan menggunakan kata kunci atau tagar #laporhendi. Dikarenakan data text bersifat tidak terstruktur sehingga memerlukan suatu analisis text untuk pemrosesan datanya. *Text mining* merupakan teknik analisis teks dalam suatu dokumen melalui kategorisasi teks, pengelompokan teks, ekstraksi konsep, dan penghapusan teks [1]. Dalam melakukan pemetaan *cluster* akan digunakan algoritma clustering dalam melakukan pembagian kelompok aduan. Algoritma clustering yang cukup populer dalam melakukan pengelompokan objek data salah satunya adalah algoritma k-means [2],[3],[4]. Pemetaan klaster didasarkan pada organisasi pemerintah daerah yang ada tersebut. Evaluasi hasil cluster dilakukan dengan menggunakan *purity* untuk mengetahui ketepatan hasil pengelompokan atau *clustering* dari data aduan.

2. Metodologi Penelitian

Pengumpulan data diperoleh dengan melakukan *crawling data* aduan masyarakat dari twitter dengan menggunakan kata kunci "*laporhendi*". Data tersebut dianalisis dengan text mining mengingat data tersebut merupakan data text. Analisis text ini diawali dengan melakukan *preprocessing* terhadap data mentah hasil *crawling data*. Tahapan *cleansing data* dilakukan untuk membersihkan data terhadap duplikasi ataupun dokumen yang sama sekali tidak bermakna, berikutnya adalah melakukan *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil *preprocessing* diperoleh dataset yang siap untuk proses pembobotan *term* dengan model TF IDF

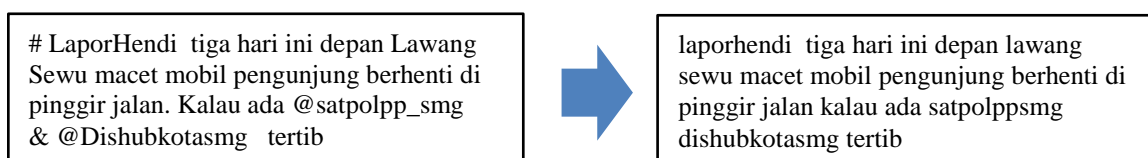
yang dinotasikan dengan nilai 0 dan 1. Barulah kemudian dilakukan klasterisasi data dengan algoritma k-means. Hasil klaster yang diperoleh akan dievaluasi dengan metode *purity* untuk mengetahui hasil *cluster* yang paling optimal.

2.1. Text Preprocessing

Pemrosesan awal untuk pengolahan data adalah untuk melakukan pembersihan terhadap data mentah yaitu data text aduan, dengan tahapan sebagai berikut :

2.1.1. Case folding

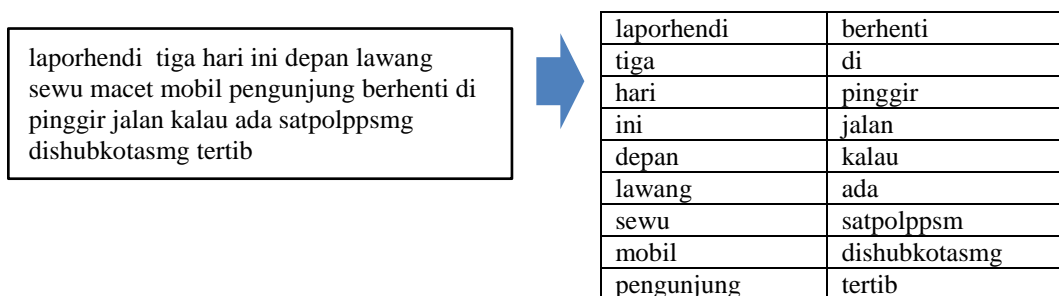
Transformasi terhadap format data text yang ada yaitu dengan mengubah huruf capital menjadi huruf kecil. Selain itu juga melakukan penghapusan terhadap konten dalam text yang bukan karakter. Berikut adalah contoh dokumen atau data tweet yang diproses.



Gambar 1. Contoh hasil proses *case folding*

2.1.2. Tokenizing

Berikutnya adalah membentuk struktur data text menjadi suatu kata untuk mewakili keseluruhan objek data text. Dimana sebelumnya data dengan struktur kalimat akan dipecah dalam kata per kata.



Gambar 2. Contoh hasil proses *tokenizing*

2.1.3. Stopword Removal

Tahap *stopword removal* ini adalah menghapus kata atau *term* yang tidak memberikan makna seperti kata hubung, misalnya “dengan”, “atau”, dan lain sebagainya. Untuk proses penghapusan kata- kata atau term adalah dengan mengikuti daftar kata yang ada pada daftar *stopword*. Untuk *stopword* yang digunakan, penelitian ini memanfaatkan library *satrawi*.

2.1.4. Stemming

Untuk mendapatkan term akhir, maka proses dilanjutkan dengan transformasi tiap kata ke dalam bentuk kata dasar, sehingga berbagai bentuk imbuhan akan dihilangkan.

2.2. Pembobotan Kata

Untuk pembentukan vector dari data text, model pembobotan *Term Frequency – Inverse Document Frequency* (TF-IDF) digunakan dalam penelitian ini dengan menggunakan persamaan berikut [5].

$$Wdt = tfdt * Idft \dots\dots\dots (1)$$

yaitu dimana :

Wdt = bobot dokumen ke $-d$ terhadap kata ke $-t$

tfdt = banyaknya kata yang dicari pada sebuah dokumen

Idft = inversed dokumen frequency ($(\log \frac{N}{df})$)

N = total dokumen

df = banyak dokumen yang mengandung kata yang dicari

2.3. Clustering

Algoritma K-Means adalah salah satu algoritma clustering tanpa adanya supervisi (*unsupervised*). Suatu metode dalam melakukan pengelompokkan dengan menggunakan sistem partisi. Prinsip yang digunakan dari algoritma K-Means ialah mengelompokkan data berdasarkan ke beberapa cluster yang telah ditentukan. Tahapan dalam penghitungan algoritma K-Means adalah sebagai berikut [6], [7], [8],[9]:

1. Tentukan banyak cluster yang akan dibentuk
2. Menentukan nilai pusat cluster awal atau titik centroid
3. Menghitung jarak tiap data yang ada ke centroid yang akan diklaster. Perhitungan jarak ini menggunakan rumus *Eucledian Distance*. Berikut persamaan *Eucledian Distance*:

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2} \dots\dots\dots (2)$$

Keterangan :

$d(x_i, \mu_i)$: Jarak antara cluster x dengan pusat cluster μ pada kata ke i

x_i : bobot data ke i pada cluster yang ingin dicari jaraknya.

μ_i : bobot kata ke i pada pusat cluster.

4. Mengelompokkan tiap data ke dalam cluster berdasarkan jarak terdekat dengan centroid
5. Menentukan nilai centroid yang baru
6. Mengulang langkah 2 hingga 3 sampai anggota tiap cluster tidak ada yang berubah
7. Jika anggota cluster tidak berubah, maka nilai rata-rata centroid pada iterasi terakhir digunakan sebagai parameter untuk klasterisasi data.

2.4 Evaluasi

Evaluasi hasil kluster dilakukan menggunakan pengukuran *purity*. Nilai rasio *purity* ini merupakan nilai dari dokumen text yang di kelompokkan dengan tepat terhadap jumlah seluruh kasus [10]. Rentan nilai rasio tersebut adalah 0 hingga 1. Jika nilainya mendekati 1 diartikan bahwa sejumlah dokumen yang dikelompokkan benar. Akan tetapi kondisi sebaliknya yaitu apabila nilai rasio mendekati 0 artinya justru semakin kecil dokumen yang dikelompokkan dengan benar. Untuk menghitung nilai *purity* menggunakan persamaan sebagai berikut [7] :

$$Purity(j) = \frac{1}{n_j} \max(n_{ij}) \dots\dots\dots (2.6)$$

Total nilai *Purity* dapat dihitung dengan rumus berikut :

$$Purity = \sum_{i=0}^j \frac{n_j}{n} Purity(j) \dots\dots\dots (2.7)$$

3. Hasil dan Pembahasan

Data sample yang diolah dalam penelitian ini disajikan dalam Tabel 1 berikut ini. Disajikan 5 sample data atau dokumen yang akan dikelompokkan ke dalam 3 cluster nantinya. Data berikut sudah melalui tahapan *preprocessing data*.

Tabel 1. Data Sample Aduan Masyarakat

D1	Case Folding	@pln_123 listrik di wil.kel.sambiroto kec.tembalang kota semarang sore ini padam. ada apa?? #laporhendi
	Cleansing	listrik di wil kel sambiroto kec tembalang kota semarang sore ini padam ada apa
	Tokenizing	listrik di wil kel sambiroto kec tembalang kota semarang sore ini padam ada apa
	Stopwords	listrik wil kel sambiroto kec tembalang kota semarang sore padam
	Stemming	listrik wilayah kelurahan sambiroto kecamatan tembalang kota semarang sore padam
D2	Case Folding	#laporhendi 3 hari ini depan lawang sewu macet mobil pengunjung berhenti di pinggir jalan. kalau ada @satpolpp_smg & @Dishbukotasmg tertib
	Cleansing	hari ini depan lawang sewu macet mobil pengunjung berhenti di pinggir jalan kalau ada tertib
	Tokenizing	hari ini depan lawang sewu macet mobil pengunjung berhenti di pinggir jalan kalau ada tertib
	Stopwords	lawang sewu macet mobil pengunjung berhenti pinggir jalan tertib
	Stemming	lawang sewu macet mobil kunjung henti pinggir jalan tertib
D3	Case Folding	@hendrarprihadi #laporhendi air pdam seminggu ini tak mengalir di perumnas sendang mulyo, gimana ini pak wali..?
	Cleansing	air pdam seminggu ini tak mengalir di perumnas sendang mulyo gimana ini pak wali
	Tokenizing	air pdam seminggu ini tak mengalir di perumnas sendang mulyo gimana ini pak wali
	Stopwords	air pdam minggu mengalir perumnas sendang mulyo
	Stemming	air pdam minggu alir perumnas sendang mulyo
D4	Case Folding	#laporhendi \pdam di wilayah manyaran kp hollywood, hari ini air mati seharian, ini knp ada giliran atau kerusakan? \ @pdamkotasmg @hendrarprihadi @p3mkotasmg
	Cleansing	pdam di wilayah manyaran kp hollywood hari ini air mati seharian ini knp ada giliran atau kerusakan
	Tokenizing	pdam di wilayah manyaran kp hollywood hari ini air mati seharian ini knp ada giliran atau kerusakan
	Stopwords	pdam wilayah manyaran kp hollywood air mati giliran kerusakan
	Stemming	pdam wilayah manyaran kp hollywood air mati gilir rusak
D5	Case Folding	#laporhendi @hendrarprihadi @Dishbukotasmg jalanan memprihatinkan,,, jati ngalah macet parah,,, mobil pada mogok.
	Cleansing	jalanan memprihatinkan jati ngalah macet parah mobil pada mogok
	Tokenizing	jalanan memprihatinkan jati ngalah macet parah mobil pada mogok
	Stopwords	jalanan memprihatinkan jati ngalah macet parah mobil mogok
	Stemming	jalan prihatin jati ngalah macet parah mobil mogok

3.1. Pembobotan Kata

Dari data pada Tabel 1, dilakukan pembobotan kata dengan TF-IDF dan diperoleh 37 term dari 5 dokumen text aduan. Hasil perhitungan pembobotan TF-IDF disajikan dalam Tabel 2 berikut ini.

Tabel 2. Pembobotan Term dari Data Sample

Term	tf						idf	W = tf*idf				
	D1	D2	D3	D4	D5	df	log(Dn/df)	D1	D2	D3	D4	D5
Listrik	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Wilayah	1	0	0	1	0	2	0,3979	0,3979	0	0	0,3979	0
Kelurahan	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
sambiroto	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
kecamatan	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
tembalang	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Kota	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Semarang	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Sore	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Padam	1	0	0	0	0	1	0,6990	0,6990	0	0	0	0
Lawang	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Sewu	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Macet	0	1	0	0	1	2	0,3979	0	0,3979	0	0	0,3979
Mobil	0	1	0	0	1	2	0,3979	0	0,3979	0	0	0,3979
Kunjung	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Henti	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Pinggir	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Jalan	0	1	0	0	1	2	0,3979	0	0,3979	0	0	0,3979
Tertib	0	1	0	0	0	1	0,6990	0	0,6990	0	0	0
Air	0	0	1	1	0	2	0,3979	0	0	0,3979	0,3979	0
Pdam	0	0	1	1	0	2	0,3979	0	0	0,3979	0,3979	0
Minggu	0	0	1	0	0	1	0,6990	0	0	0,6990	0	0
Alir	0	0	1	0	0	1	0,6990	0	0	0,6990	0	0
Perumnas	0	0	1	0	0	1	0,6990	0	0	0,6990	0	0
Sendang	0	0	1	0	0	1	0,6990	0	0	0,6990	0	0
Mulyo	0	0	1	0	0	1	0,6990	0	0	0,6990	0	0
Manyaran	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0
Kp	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0
hollywood	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0
Mati	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0
Gilir	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0
Rusak	0	0	0	1	0	1	0,6990	0	0	0	0,6990	0

Prihatin	0	0	0	0	1	1	0,6990	0	0	0	0	0,6990
Jati	0	0	0	0	1	1	0,6990	0	0	0	0	0,6990
Ngaleh	0	0	0	0	1	1	0,6990	0	0	0	0	0,6990
Parah	0	0	0	0	1	1	0,6990	0	0	0	0	0,6990
Mogok	0	0	0	0	1	1	0,6990	0	0	0	0	0,6990

3.2. Clustering dengan Algoritma K-means

Sebagai langkah awal dalam clustering yaitu perlu ditentukan titik pusat awal atau titik centroid awal. Tabel 3 berikut menjadi centroid awal yang dipilih secara acak yang diambil dari dokumen *term* :

Tabel 3. Titik centroid

Nilai Titik Awal Cluster (<i>centroid</i>)	
C1	= D5
C2	= D1
C3	= D3

Selanjutnya dilakukan perhitungan jarak tiap data dengan titik pusat klaster awal yang sudah ditentukan. Dengan menggunakan fungsi *euclidian distance* dihitung jarak tiap data tersebut ke masing-masing titik pusat dengan hasil disajikan pada Tabel 4 sebagai berikut :

Tabel 4. Hasil perhitungan jarak euclidian distance iterasi 1

Centroid	Jarak ke centroid				
	D1	D2	D3	D4	D5
C1	2,7337	2,3182	2,3827	2,5148	0
C2	0	2,8217	2,7046	2,7650	2,7337
C3	2,7046	2,4831	0	2,3521	2,3827

Setelah diperoleh jarak antar centroid dengan data maka ditentukan jarak yang paling dekat dari setiap data dengan centroid. Yang nantinya jarak terdekat merupakan kelompok data pada cluster tersebut. Berikut kelompok anggota cluster pada iterasi 1 :

Tabel 5. Pengelompokan anggota cluster iterasi 1

Pengelompokan anggota <i>cluster</i> Iterasi 1					
	D1	D2	D3	D4	D5
C1	2,7337	2,3182	2,3827	2,5148	0
C2	0	2,8217	2,7046	2,7650	2,7337
C3	2,7046	2,4831	0	2,3521	2,3827

Berikutnya dilakukan perhitungan jarak tiap dokumen ke titik centroid baru dan menghasilkan anggota cluster seperti pada Tabel 6 berikut.

Tabel 6. Hasil jarak tiap data pada iterasi 2

Pengelompokan anggota <i>cluster</i> Iterasi 2					
	D1	D2	D3	D4	D5
C1	2,5247	1,1591	2,1397	2,2858	1,1591
C2	0	2,8217	2,7046	2,7650	2,7337
C3	2,4692	2,5297	1,1761	1,1761	2,1489

Pada iterasi kedua, didapati bahwa anggota kluster tidak mengalami perubahan sehingga proses perhitungan jarak berhenti dan diperoleh hasil ahir anggota kluster sebagai berikut.

3.3. Evaluasi Hasil Kluster

Sedangkan untuk menghitung nilai *purity* setiap cluster pada hasil perhitungan data diatas dengan jumlah 5 Data dan 3 Cluster sebagai berikut :

Tabel 7. Hasil evaluasi cluster dengan *purity*

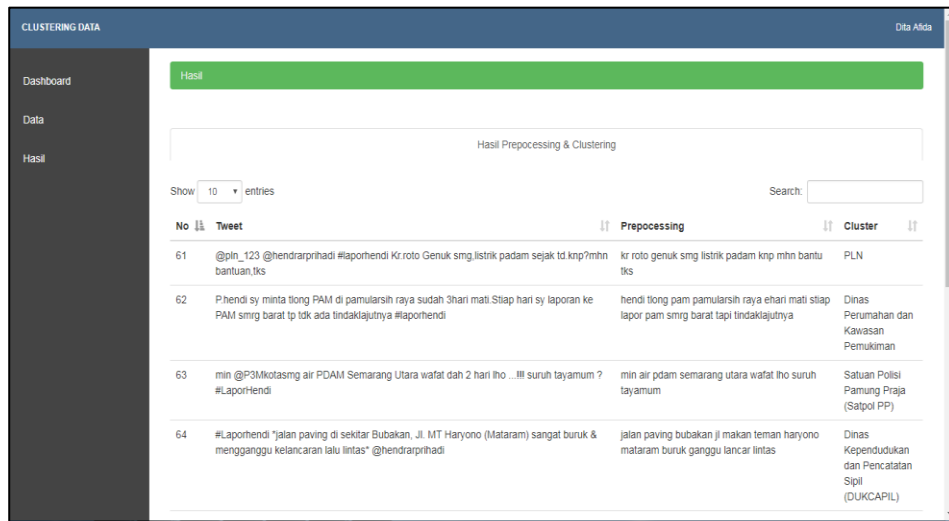
Cluster	Jumlah	PLN	PDAM	Dishub	Purity
C1	2	0	0	2	1
C2	1	1	0	0	1
C3	2	0	2	0	1
Total	5				
Purity					1

Dari evaluasi hasil cluster tersebut diperoleh nilai *purity* untuk 3 cluster dengan 5 data *sample* adalah sebesar 1. Sehingga hasil evaluasi pada data *sample* yang digunakan memiliki hasil yang baik karena bernilai 1, artinya setiap data masuk ke cluster yang sesuai dengan tepat. Untuk selanjutnya model cluster akan dibangun dari sejumlah data yang sudah diperoleh dari *proses crawling* data pada twitter.

3.4 Dashboard Aplikasi Klustering Aduan Masyarakat

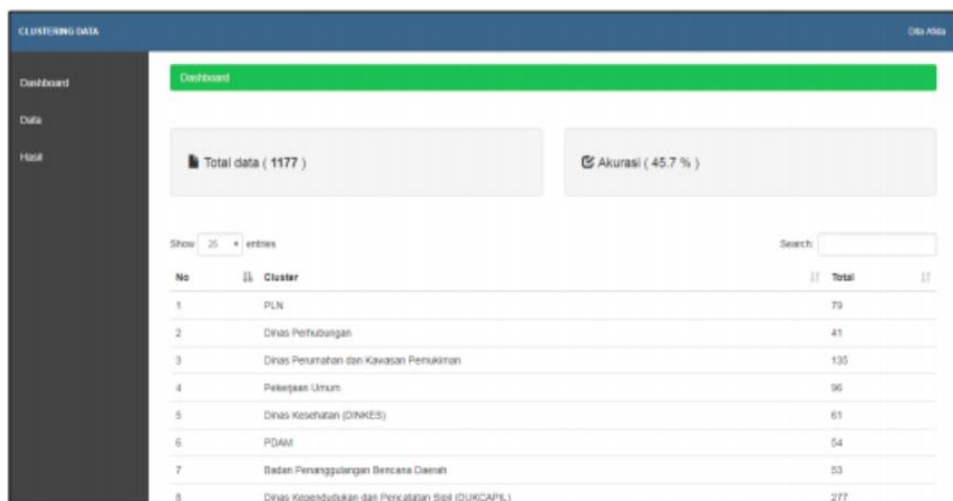
Model dari hasil analisis cluster untuk data aduan menggunakan algoritma *k-means* selanjutnya diimplementasikan ke dalam prototype dari aplikasi analisis text aduan masyarakat. Hasil dari clustering data yang telah diinputkan yaitu terbagi menjadi 11 OPD yang ada dan menjadi acuan banyaknya jumlah *cluster* yang dibentuk. Aplikasi akan menampilkan data aduan yang masuk kemudian mengkategorikan setiap aduan tersebut ke masing- masing unit atau dalam hal ini OPD yang sesuai. Aplikasi analisis text aduan masyarakat yang dikembangkan disajikan dalam Gambar 3 berikut ini. Secara keseluruhan, data yang dikumpulkan dan dianalisis adalah sejumlah 1171 data aduan.

Pada Gambar 3, aplikasi menampilkan data aduan yang masuk dan kemudian memberikan keterangan hasil clusternya. Sehingga aplikasi analisis text aduan masyarakat ini akan secara otomatis mengelompokkan cluster setiap aduan yang masuk. Petugas terkait tidak lagi kesulitan dalam membaca maupun mengelompokkan aduan yang masuk karena sistem sudah secara langsung mengelompokkan. Harapannya dari aplikasi yang dirancang ini maka dapat membantu petugas mempercepat proses tindak lanjut aduan ke OPD terkait. Yang pada akhirnya tentu dapat meningkatkan layanan pemerintah kepada masyarakat.



Gambar 3. Dashboard hasil klastering

Pada Gambar 4 dibawah ini adalah merupakan tampilan dashboard untuk keseluruhan data yang telah di kelompokkan ke dalam berbagai cluster atau OPD yang ada. Nampak bahwa aplikasi memberikan informasi total aduan masyarakat yang masuk untuk setiap cluster.



Gambar 4. Dashboard total aduan yang dikelompokkan ke setiap cluster

4. Kesimpulan

Dari hasil analisis text mining yang telah dilakukan pada data text aduan masyarakat yang bersumber dari twitter. Dapat disimpulkan bahwa hasil pengelompokkan (*cluster*) data dengan algoritma k-means mendapatkan hasil yang bagus. Hal ini didapat dari nilai purity yang mana nilainya adalah 1 untuk tiap cluster yang ada. Oleh karena itu dalam melakukan klasterisasi text aduan masyarakat dapat memanfaatkan algoritma k-means. Dan model cluster yang diterapkan dalam aplikasi analisis text aduan masyarakat dapat membantu pihak terkait dalam proses otomatisasi pengelompokkan aduan masyarakat ke dinas terkait. Sehingga proses tindak lanjut aduan menjadi lebih efektif.

References

- [1] M. Afzali and S. Kumar, "Text Document Clustering : Issues and Challenges," *2019 Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput.*, pp. 263–268, 2019.
- [2] X. et al Wu, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, pp. 1–37, 2008.
- [3] S. Karyadi and H. Yasin, "Analisis Kecenderungan Informasi Dengan Menggunakan Metode Text Mining," *J. Gaussian*, vol. 5, pp. 763–770, 2016.
- [4] D. S. Indraloka, B. Santosa, D. Matematika, F. Matematika, P. Alam, I. Teknologi, and S. Nopember, "Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia," *J. Sains dan Seni*, vol. 6, no. 2, pp. 6–11, 2017.
- [5] R. Melita, V. Amrizal, H. B. Suseno, T. Dirjam, P. Studi, T. Informatika, and F. Sains, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus : Syarah Umdatil Ahkam)," *J. Tek. Inform.*, vol. 11, no. 2, 2018.
- [6] R. K. Dinata, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *Informatics J.*, vol. 5, no. 1, 2020.
- [7] J. Teknik, I. Fik, and J. N. N. Semarang-, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," Universitas Dian Nuswantoro, 2015.
- [8] S. Defiyanti, M. Jajuli, T. Informatika, F. Ilmu, K. Universitas, and S. Karawang, "Implementasi Algoritma K-Means Dalam," *J. Ilm. Inf. Terap.*, vol. I, no. 2, pp. 62–68, 2015.
- [9] A. K. Wardhani, "Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, pp. 30–37, 2016.
- [10] K. R. Prilianti and K. Kunci, "Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering," *J. Cybermatika*, vol. 2, no. 1, pp. 1–6, 2014.