

## ANALISIS TOPIK EKONOMI DENGAN ALGORITMA *K-MEANS* PADA MEDIA *ONLINE* ERA PANDEMI COVID-19 DI SULAWESI TENGGARA

Muhammad A Ayub<sup>1</sup>

<sup>1</sup>Badan Pusat Statistik Kabupaten Konawe  
Email: <sup>1</sup>Muhammad.ayub@bps.go.id

(Naskah masuk: 21 Juni 2021, diterima untuk diterbitkan: 22 Juli 2021)

### Abstrak

Pandemi Covid-19 sejak maret 2020 hingga saat ini telah banyak memberikan dampak terhadap berbagai sektor, termasuk diantaranya adalah sektor ekonomi. Berbagai isu pun semakin berkembang terkait perekonomian di era pandemi Covid-19 khususnya di Sulawesi Tenggara. Adanya media *online* sebagai alat utama untuk mengetahui perkembangan informasi baik itu berupa isu atau fakta yang terjadi masih menyulitkan masyarakat karena banyaknya pemberitaan dan keberadaan media *online*. Kajian ini bertujuan untuk menggambarkan tren topik ekonomi dan bisnis pada pemberitaan media *online* di Sulawesi Tenggara era pandemi Covid-19. Penentuan tren topik ini dimulai dengan pengumpulan data menggunakan teknik web *scraping* judul berita kategori ekonomi dan bisnis pada 4 media *online*. Judul-judul berita kemudian dibentuk menjadi beberapa kluster dengan bantuan visualisasi *wordcloud* dengan terlebih dahulu melakukan *preprocessing* data. Hasil penelitian menunjukkan bahwa dengan menggunakan algoritma *K-Means* diperoleh 3 kluster dengan nilai WSS sekitar 4,4 % dan koefisien *Sillhouete* sekitar 0,1. Adapun tren topik ekonomi pada media *online* di Sulawesi Tenggara era Pandemi Covid-19 yaitu bantuan sosial, pelaku dan tindakan ekonomi, serta regulasi keuangan.

**Kata kunci:** Tren Topik, Kluster, *Web Scraping*, *K-Means*

## ANALYSIS OF ECONOMIC TOPIC WITH *K-MEANS* ALGORITHM ON *ONLINE* MEDIA OF THE COVID-19 ERA IN SULAWESI TENGGARA

### Abstract

The Covid-19 pandemic since March 2020 has given many impacts in various sectors, including the economic sector. Various issues are also growing related to the economy in the era of the Covid-19 pandemic, especially in Southeast Sulawesi. The existence of online media as the main tool to find out the information whether in the form of issues or facts is still difficult for people to understand because of the large amount of news and the existence of online media. This study aims to describe trends in economic and business topics on online media coverage in Southeast Sulawesi during the Covid-19 pandemic era. Determining the trend of this topic begins with data collection using web scraping techniques for news headlines in the economic and business categories on 4 online media. The news titles are then formed into several clusters with the help of wordcloud visualization by first doing data preprocessing. The results showed that using the *K-Means* algorithm obtained 3 clusters with a WSS value of about 4.4% and the *Sillhouete* coefficient of about 0.1. The resulting economic topic trends on online media of the covid-19 era in Southeast Sulawesi are social assistance, economic actors and actions, and financial regulation.

**Keywords:** Trending Topics, Cluster, *Web Scraping*, *K-Means*

### 1. PENDAHULUAN

Pencarian informasi sesuai kebutuhan di jaman digitalisasi saat ini semakin sulit dikarenakan tingginya tingkat ketersediaan informasi yang ada [1]. Hal ini ditambah dengan tingginya peningkatan pengguna internet yang secara langsung memacu pertumbuhan jumlah media *online* yang ada dan perlahan mengubah perilaku masyarakat dalam mengakses berita dari media cetak ke media *online* [2]. Dengan begitu pola pikir dan pengetahuan

masyarakat akan suatu informasi sangat dipengaruhi oleh pemberitaan pada media *online*.

Pandemi Covid-19 sejak maret 2020 hingga saat ini telah banyak memberikan dampak terhadap berbagai sektor, termasuk diantaranya adalah sektor ekonomi. Badan Pusat Statistik (BPS) Provinsi Sulawesi Tenggara mencatat ekonomi Sulawesi Tenggara tahun 2020 mengalami kontraksi sebesar 0,65 persen dibanding capaian tahun 2019 yang tumbuh sebesar 6,50 persen [3]. Berbagai isu pun

semakin berkembang terkait perekonomian di era pandemi Covid-19 khususnya di Sulawesi Tenggara. Media *online* sebagai salah satu media informasi di era teknologi saat ini dapat menjadi alat utama untuk mengetahui perkembangan informasi baik itu berupa isu atau fakta yang terjadi [4]. Hanya saja banyaknya pemberitaan dan keberadaan media *online* sedikit menyulitkan masyarakat untuk mengetahui tren informasi yang ada khususnya informasi ekonomi dan bisnis di era pandemi Covid-19.

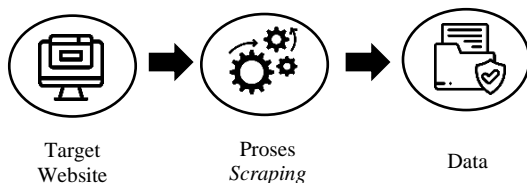
Tujuan dari tulisan ini adalah untuk mengetahui tren topik ekonomi dan bisnis pada media *online* di Sulawesi Tenggara selama masa pandemi Covid-19. Hal ini mengingat sektor ekonomi dan bisnis sangat terdampak akibat pandemi Covid-19 [5]. Dengan mengetahui tren topik ekonomi dan bisnis maka masyarakat akan mendapatkan perkembangan tren informasi sektor ekonomi dan bisnis serta tidak menutup kemungkinan untuk dijadikan gambaran untuk melakukan kajian pada bidang ekonomi secara lebih spesifik khususnya pada era pandemi Covid-19.

## 2. METODE PENELITIAN

Unit analisis dalam kajian ini adalah studi kasus pada empat media *online* di Provinsi Sulawesi Tenggara yaitu *kendaripos.com*, *mediakendari.com*, *zonasultra.com*, dan *beritakotakendari.com*. Adapun data yang digunakan adalah judul berita dengan kategori ekonomi dan bisnis pada masing-masing media *online* tersebut periode 1 April 2020 s.d 31 Maret 2021 (Era Pandemi Covid di Sulawesi Tenggara).

### 2.1. Web Scraping

*Web Scraping* adalah teknik pengambilan data dan informasi pada suatu halaman website [6]. Pengambilan informasi diawali dengan mempelajari dokumen HTML dari website yang menjadi target, lalu menyusun algoritma untuk pengambilan informasi yang dibutuhkan secara otomatis [7]. Data yang diperoleh dari hasil *Web Scraping* kemudian disimpan untuk selanjutnya dilakukan analisis sesuai kebutuhan. Langkah-langkah proses *Web Scraping* dapat dilihat pada gambar 1.

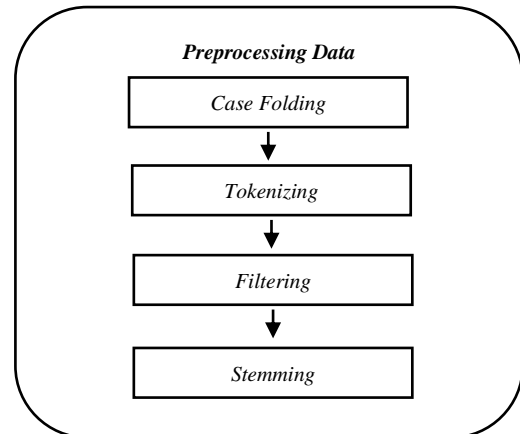


Gambar 1. Tahapan Web Scraping

### 2.2. Preprocessing Data

*Preprocessing* data adalah tahapan awal pengolahan *text mining* yang bertujuan untuk mentransformasi data agar lebih bermakna [8].

Langkah-langkah *Preprocessing* data dapat dilihat pada gambar 2.



Gambar 2. Langkah-langkah *Preprocessing* Data

Pada tahapan awal yaitu *Case Folding* seluruh data teks yang ada diubah menjadi huruf kecil. Selanjutnya ke tahap *Tokenizing* atau pemisahan kata-kata menjadi *list* serta menghilangkan beberapa karakter seperti tanda baca dan sebagainya. Kemudian tahapan ketiga *Filtering* dimana menghapus kata-kata yang tidak diperlukan dengan menggunakan algoritma *stopword* atau *stoplist* yang sudah disesuaikan dengan kebutuhan kajian ini. Tahap terakhir yaitu *Stemming* merupakan tahap mencari kata dasar. Dalam kajian ini tahap *stemming* tidak dilakukan karena dalam teks bahasa Indonesia justru mengurangi akurasi [9].

### 2.3. Term Weighting

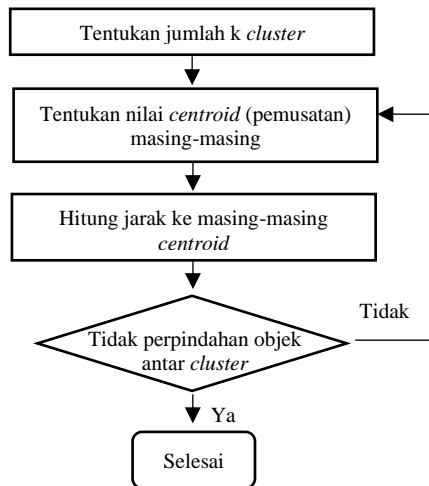
*Term Weighting* adalah jenis metode untuk menghitung pembobotan kemunculan suatu istilah pada suatu dokumen [9]. Dalam *Term Weighting*, metode yang digunakan dalam melakukan pembobotan dikenal dengan dua hal yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) atau lebih dikenal dengan istilah metode TF-IDF [5]. *Term Frequency* (TF) menyatakan frekuensi sebuah kata dalam dokumen sementara *Inverse Document Frequency* (IDF) menyatakan kebalikan dari kemunculan sebuah kata pada beberapa dokumen yang berbeda [10]. Adapun pembobotan metode TF-IDF ( $w_{ij}$ ) merupakan perkalian dari nilai TF dan IDF yang dapat dilihat pada persamaan (1),  $tf_{ij}$  adalah frekuensi kata  $i$  dalam dokumen ke  $j$ ,  $df_i$  adalah frekuensi dokumen dalam kata  $i$ , dan  $N$  adalah jumlah dokumen.

$$w_{ij} = tf_{ij} \times \log_2 \left( \frac{N}{df_i} \right) \quad (1)$$

### 2.4. K-Means Cluster

*Clustering* adalah salah satu metode pengelompokan dimana pola pembentukan suatu *cluster* dengan *cluster* lainnya memiliki kesamaan

ciri. Dalam membentuk *cluster* ada banyak metode, namun dalam kajian ini menggunakan algoritma *K-Means* untuk data teks. Pemilihan metode *K-Means* dalam pembentukan *cluster* tidak terlepas karena merupakan salah satu yang terpopuler [11]. Selain itu, penggunaan algoritma *K-Means* untuk pengolahan data dalam jumlah yang besar sangat efektif [12]. Adapun langkah-langkah algoritma *K-Means* dapat dilihat pada gambar 3 [13].



Gambar 3. Algoritma K-Means

Dalam penentuan jumlah *k cluster* yang akan dibentuk terdapat banyak metode. Dalam kajian ini akan digunakan beberapa metode yaitu Metode *Elbow*, *Silhouette* dan *Within Cluster Sum of Squares (WSS)*. Metode *Elbow* memberikan informasi jumlah kluster terbaik melalui perbandingan dua kluster terdekat dengan nilai *Sum of Square Error (SSE)* yang mengalami penurunan terbesar [14]. Metode *Elbow* ini merupakan metode paling umum untuk menentukan jumlah *cluster* optimal [15]. Metode *Silhouette Coefficient* merupakan metode pengukuran kekuatan dan kualitas sebuah *cluster* dengan memperhatikan kedekatan antar objek suatu *cluster* yang sama dan seberapa jauh sebuah *cluster* dengan *cluster* lainnya [16]. *Within Cluster Sum of Squares (WSS)* adalah metode pengukuran keragaman intra kluster [8].

### 3. HASIL DAN PEMBAHASAN

Hasil kajian ini diawali implementasi teknik *web scraping* mengenai judul berita kategori ekonomi dan bisnis pada beberapa media *online*. *Scraping* dilakukan dengan menggunakan Bahasa pemrograman *javascript* yang hasilnya diexport kedalam bentuk file *CSV*. Dari hasil *Scraping* diperoleh 1593 judul berita kategori ekonomi dan bisnis. Dengan rinciannya dapat dilihat pada tabel 1.

Tabel 1. Ringkasan Data Hasil *Scraping*

Nama Media Online	Jumlah Judul Berita Kategori Ekonomi dan Bisnis
Kendari Pos	584
Media Kendari	344
Zona Sultra	324
Berita Kota Kendari	341

Judul-judul yang diperoleh tersebut kemudian dilakukan *preprocessing* data teks agar dapat lebih bermakna saat dilakukan analisis lanjutan. Berikut ilustrasi tahapan *preprocessing* data teks pada salah satu judul yang dapat dilihat pada tabel 2.

Tabel 2. Ilustrasi tahapan *preprocessing*

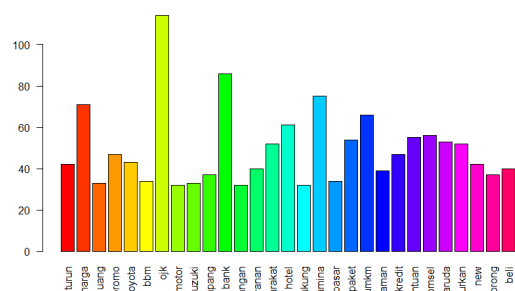
Tahapan <i>Preprocessing</i>	Hasil
Teks Awal	FKIJK & OJK Sultra Salurkan 250 Paket Bansos Kepada Masyarakat Kurang Mampu!
<i>Case Folding</i>	fkijk & ojk sultra salurkan 250 paket bansos kepada masyarakat kurang mampu!
<i>Tokenizing</i>	fkijk ojk sultra salurkan 250 paket bansos kepada masyarakat kurang mampu
<i>Filtering</i>	fkijk ojk sultra salurkan masyarakat kurang mampu

Tabel 2 menunjukkan ilustrasi tahapan *preprocessing* data teks pada salah satu judul berita yang akan dikaji. Hasil *preprocessing* data teks ini mampu menghapus jumlah kata tidak bermakna sebesar 8,53 % dari keseluruhan kata yang ada pada data awal. Hal ini tentunya sangat membantu dalam melakukan kajian data teks pada analisis lanjutan.

Kumpulan kata-kata pada semua judul berita kemudian dilakukan pembobotan kata dengan metode *TF-IDF* sebagai langkah awal sebelum melakukan klusterisasi topik. Ringkasan matriks *TF-IDF* yang terbentuk dapat dilihat pada tabel 3.

Tabel 3. Ringkasan Matriks *TF-IDF*

Term	D1	D2	D3	...	D1593
turun	1,57	0	0	...	0
harga	0	0	1,35	...	0
kesejahteraan	0	3,20	0	...	0
prediksi	0	0	2,50	...	0
naik	0	0	2,30	...	0
ekspor	1,90	0	0	...	0
layanan	0	0	0	...	0
solusi	0	2,42	0	...	0
....	...	...	...	...	...
lawan	0	0	0	...	3,20

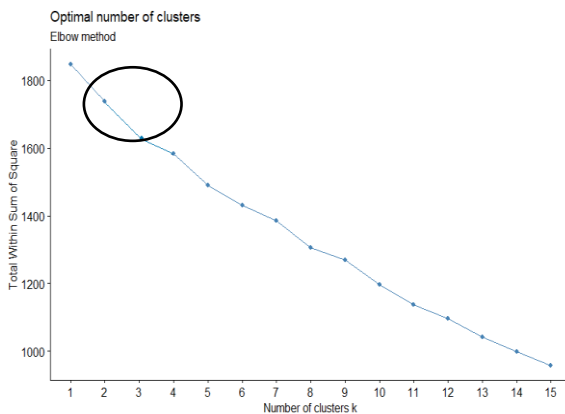


Gambar 3. Diagram Batang Kemunculan Term dari 1593 judul

Berdasarkan gambar 3 terlihat term yang paling banyak muncul pada semua dokumen dimana memperlihatkan bahwa kata “ojk” muncul paling sering dengan frekuensi lebih dari 100 kali. Selain itu beberapa kata lainnya yang sering muncul diantaranya “bank”, “pertamina”, “harga”, “umkm”, dan lain sebagainya.

Selanjutnya pada tahap klusterisasi dilakukan dengan menggunakan matriks TDF-IF yang telah dibuat. Langkah awal dilakukan dengan menentukan jumlah k kluster yang akan dibentuk. Kluster yang dibentuk diharapkan paling sedikit 3 agar penentuan topik utama pada tiap kluster dapat lebih baik. Berdasarkan hasil pengamatan dengan metode *elbow* pada gambar 4 serta Nilai *Within Cluster Sum of Squares* (WSS) dan Koefisien *Silhouette* pada tabel 4 terlihat bahwa kombinasi kluster optimal yang dibentuk adalah 3 kluster.

Gambar 4 dengan grafik *Elbow* terlihat bahwa k=3 paling optimal, dimana perubahan *Sum Square Error* terlihat paling besar. Hal ini didukung pula dengan nilai WSS terkecil dan Koefisien *Silhouette* terbesar dibanding dengan kluster lainnya.



Gambar 4. Grafik *elbow*

kluster (k)	WSS	Si
3	4,40 %	0,10
4	8,20 %	0,08
5	13,80 %	0,07
6	16,60 %	0,05

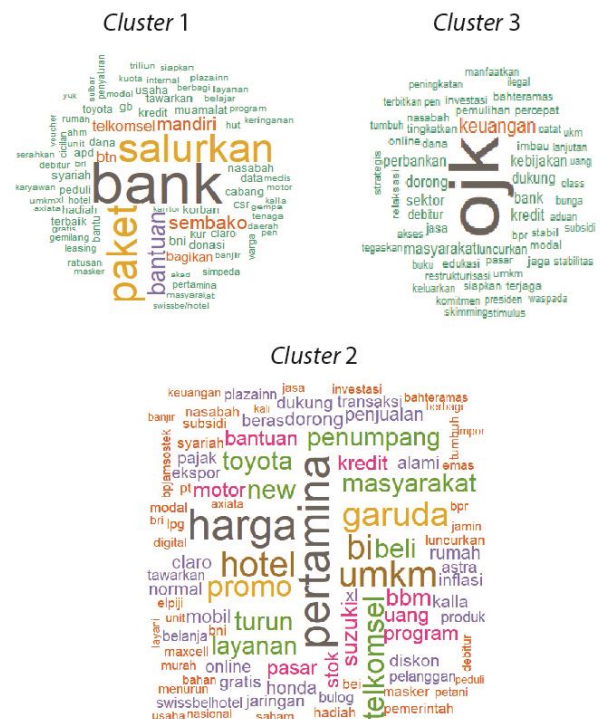
Tahap klusterisasi dengan algoritme *K-Means* dengan k=3, diperoleh informasi sebagai berikut :

Kluster	Jumlah Anggota Kluster
Kluster 1	161
Kluster 2	1318
Kluster 3	114

Dari tabel 5 terlihat bahwa jumlah anggota antar kluster sangat timpang, dimana kluster dua memiliki anggota terbesar sebanyak 1318 anggota. Sementara untuk kluster satu dan tiga jumlah anggota hanya berkisar seratusan anggota.

Selanjutnya, untuk mengetahui tren topik dilakukan visualisasi kata-kata pembentuk judul dalam masing-masing kluster yang terbentuk dengan menggunakan *wordcloud*. Adapun gambarannya dapat dilihat pada gambar 5.

Gambar 5 terlihat bahwa pada *cluster* satu kata-kata yang paling sering muncul yaitu bank, saluran, paket, bantuan, dan sembako, dimana saat kita mengkaji secara dalam pada judul-judul berita topik ekonomi yang ada. kata-kata ini mengarah pada topik yang lebih umum yaitu mengenai bantuan sosial. Selanjutnya pada *cluster* dua terlihat bahwa dominan kata tidak terlalu terlihat langsung mengingat besarnya anggota pada *cluster* dan frekuensi kemunculan kata seimbang. Namun jika diperhatikan dengan cermat terdapat kata-kata pertamina, harga, hotel, umkm, bi, promo, garuda, beli, *new*, layanan, layanan, masyarakat, telkomsel, toyota dan penumpang. Kata-kata tersebut mewakili topik umum pelaku dan tindakan ekonomi. Sementara pada *cluster* ketiga kata dominannya yaitu ojk dan keuangan. Kedua kata ini mewakili topik umum tentang regulasi keuangan.



Gambar 5. *Wordcloud* Pada Masing-masing *Cluster*

Gambar ketiga *cluster* terlihat bahwa 3 topik umum yang kemudian nantinya dapat dikembangkan terlihat bahwa tren topik ekonomi pada media *online* di Sulawesi Tenggara era pandemi Covid-19 yang dianalisis melalui judul dan kata-kata pembentuk judulnya yaitu bantuan sosial, pelaku dan tindakan ekonomi serta regulasi keuangan. Pemberitaan topik bantuan sosial di era pandemi sangat banyak dimana terlihat bantuan-bantuan tersebut dalam upaya

menstimulus perekonomian yang sedang turun. Selain bantuan sosial, aktivitas para pelaku ekonomi masih jadi salah satu topik utama. Beberapa pelaku ekonomi seperti umkm, Pertamina, Garuda dan lain sebagainya gencar melakukan kegiatan yang tujuannya juga meningkatkan perekonomian baik dengan melakukan promo, memberikan layanan baru, atau mengatur harga. Topik selanjutnya yaitu mengenai regulasi keuangan dimana OJK selaku badan otoritas keuangan melakukan berbagai macam tindakan dan kebijakan untuk pemulihan ekonomi, stimulus serta lain sebagainya.

#### 4. KESIMPULAN

Kajian tren topik ekonomi pada media online di Sulawesi Tenggara era pandemi Covid-19 dengan algoritma *K-means* diperoleh 3 kluster topik dengan nilai WSS terendah sekitar 4,4 % dan koefisien *Silhouette* tertinggi sekitar 0,1. Pada kluster pertama topik dominan seputar bantuan sosial, kluster kedua seputar pelaku dan tindakan ekonomi, dan pada kluster ketiga seputar regulasi keuangan. Hasil kajian berupa tren topik ini dapat dilakukan analisis lanjutan secara spesifik oleh para ahli ekonomi dan peneliti lainnya untuk memotret atau menggambarkan perekonomian di era Pandemi Covid-19 khususnya di Sulawesi Tenggara.

#### 5. DAFTAR PUSTAKA

- [1] U.T. Setihohatmo, S. Rachmat., T. Susilawati and Y. Rahman. 2020. "Analisis Metoda *Latent Dirichlet Allocation* untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik", *Prosiding The 11<sup>th</sup> Industrial Research Workshop and National Seminar Bandung* 26-27 Agustus 2020, pp. 402-408. Doi: 10.35313/irwns.v11i1.2040
- [2] R. Siringoringo, Jamaluddin and R. Perangin-Angin. 2020. "Pemodelan Topik Berita Menggunakan *Latent Dirichlet Allocation* dan *K-Means Clustering*", *Jurnal Informatika Kaputama*, vol. 4(2), pp. 216-222. Doi: 10.1234/jik.v4i2.263
- [3] Badan Pusat Statistik Provinsi Sulawesi Tenggara. 2021. "Pertumbuhan Ekonomi Sulawesi Tenggara Triwulan IV 2020", *Berita Resmi Statistik*, No. 14/02/74/Th.XIV. Tersedia melalui: <<https://sultra.bps.go.id/pressrelease/2021/02/05/897/ekonomi-sulawesi-tenggara-2020-kontraksi-0-65-persen.html>> [Diakses 30 April 2021]
- [4] A. Ready. 2016. "Penggunaan Media Online Sebagai Sumber Informasi Akademik Mahasiswa Ilmu Komunikasi Fakultas Ilmu Sosial dan Ilmu Politik Universitas Riau". *Jurnal Online Mahasiswa Universitas Riau*, Vol 3. No. 1 Februari 2016. Tersedia melalui: <<https://jom.unri.ac.id/index.php/JOMFSIP/article/view/8438>> [Diakses 21 April 2021]
- [5] W. Hadiwardoyo. 2020. "Kerugian Ekonomi Nasional Akibat Pandemi Covid-19", *Journal of Business & Entrepreneurship Universitas Muhammadiyah Jakarta*, Vol 2. No. 2, pp.83-92. Doi: 10.24853/baskara.2.2.83-92
- [6] Y. Sahria. 2020. "Implementasi Teknik *Web Scraping* pada jurnal SINTA untuk analisis topik penelitian kesehatan Indonesia", *The 11<sup>th</sup> University Research Colloquium 2020*, pp. 297-306. Tersedia melalui: <<http://repository.urecol.org/index.php/proceeding/article/view/1079>> [Diakses 1 April 2021]
- [7] M. R. Ma'arif. 2016. "Integrasi laman web tentang pariwisata daerah istimewa Yogyakarta memanfaatkan teknologi *web scraping* dan *text mining*", *Jurnal Teknomika*, vol 9(2), pp. 71-80. Tersedia melalui: <<http://ejournal.unjaya.ac.id/index.php/teknomatika/article/view/461>> [Diakses 20 April 2021]
- [8] W. Hardi, W. A. Kusuma and S. Basuki. 2019. "Pengelompokan Topik Dokumen Berbasis *Text Mining* Dengan Algoritme *K-means*", *Jurnal Visi Pustaka*, Vol. 21 No.1, pp. 67-76. Doi: 10.37014/visi%20pustaka.v21i1.77.
- [9] A. F Hidayatullah and M.R Ma'arif. 2016. "Penerapan *Text Mining* dalam Klasifikasi Judul Skripsi. Seminar Nasional Aplikasi Teknologi Informasi", *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2016* Yogyakarta, pp. 33-36. Tersedia melalui: <<http://journal.uui.ac.id/Snati/article/view/6232>> [Diakses 15 April 2021]
- [10] N.G. Yudiarta, M. Sudarma and W. G. Ariastina. 2018. "Penerapan Metode Cluster Text Mining untuk pengelompokan berita pada unstructured textual data", *Majalah Ilmiah Teknologi Elektro*, Vol. 17 No. 3, pp. 339-343. Doi: 10.24843/MITE.2018.v17i03.P06
- [11] R. Adhitama, A. Burhanuddin and R. Ananda. 2020. "Penentuan Jumlah Cluster Ideal SMK di Jawa Tengah dengan Metode *X-Means Clustering* dan *K-Means Clustering*", *JIKO (Jurnal Informatika dan Komputer)*, Vol 3 No.1, pp. 1-5. Doi: 10.33387/jiko.v3i3.1635
- [12] W. Saputro, M. R. Pahlevi and A. Wibowo. 2020. "Analisis Algoritma *K-Means* Untuk Klasterisasi Tindak Pidana Korupsi di Wilayah Hukum Indonesia", *JIKO (Jurnal Informatika dan Komputer)*, Vol 3 No.3, pp. 137-142. Doi: 10.33387/jiko.v3i3.1960

- [13] E. Yulian. 2018. “Text Mining dengan *K-means Clustering* pada Tema LGBT dalam arsip *tweet* Masyarakat Kota Bandung”., *Jurnal Matematika “MANTIK”*, Vol. 4 No.1, pp. 53-58. Doi: 10.15642/mantik.2018.4.1.53-58
- [14] D. A. I. C. Dewi and D. A. K. Pramita. 2019. “Analisis Perbandingan Metode *Elbow* dan *Sillhouette* pada Algoritma *Clustering K-Medoids* dalam pengelompokan produksi Kerajinan Bali”., *Jurnal Matrix*, Vol 9 No.3, pp. 102-109. Doi: 10.31940/matrix.v9i3.1662
- [15] C. Shi, B. Wei, S. Wei, *et al.* 2021. “A Quantitative Discriminant Method of Elbow Point for The Optimal Number of Clusters in Clustering Algorithm”., *Journal on Wireless Communications and Networking*, 2021:31, Doi: 10.1186/s13638-021-01910-w
- [16] M. S. Hudin, M. A. Fauzi and S. Adinugroho. 2018. “Implementasi Metode *Text Mining* dan *K-Means Clustering* untuk pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya)”., *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol 2 No.11, pp. 5518-5524, Tersedia melalui: <<https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3332>> [Diakses 19 April 2021]