

## ***Data Mining for Determining The Best Cluster Of Student Instagram Account As New Student Admission Influencer***

Data Mining untuk Menentukan Klaster Terbaik dari Akun Instagram Mahasiswa Sebagai Influencer Penerimaan Mahasiswa Baru

**Ahmad Irfan Abdullah<sup>1</sup>, Adri Priadana<sup>2</sup>, Muhajir<sup>3</sup>, Syahrir Nawir Nur<sup>4</sup>**

<sup>1,3,4</sup> Desain Grafis, Politeknik Negeri Media Kreatif, Indonesia

<sup>2</sup> Informatika, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

<sup>1</sup>irfan@polimedia.ac.id, <sup>2\*</sup>dana@unjaya.ac.id, <sup>3</sup>muhajir@polimedia.ac.id,

<sup>4</sup>syahrirawir@polimedia.ac.id

### ***Informasi Artikel***

*Received: May 2021*

*Revised: June 2021*

*Accepted: June 2021*

*Published: August 2021*

### ***Abstract***

***Purpose:*** This study aims to apply the web data extraction method to extract student Instagram account data and the K-Means data mining method to perform clustering automatically to determine the best cluster of students' Instagram accounts as influencers for new student admissions.

***Design/methodology/approach:*** This study implemented the web data extraction method to extract student Instagram account data. This study also implemented a data mining method called K-Means to cluster data and the Silhouette Coefficient method to determine the best number of clusters.

***Findings/result:*** This study has succeeded in determining the seven best student accounts from 100 accounts that can be used as influencers for new student admissions with the highest Silhouette Score for the number of influencers selected between 5-10, which is 0.608 of the 22 clusters.

***Originality/value/state of the art:*** Research related to the determination of the best cluster of students' Instagram accounts as new student admissions influencers using web data extraction and K-Means has never been done in previous studies.

### ***Abstrak***

***Tujuan:*** Peneliti ini bertujuan untuk menerapkan metode web data extraction untuk mengekstraksi data akun Instagram mahasiswa dan metode data mining yaitu K-Means untuk melakukan klasterisasi secara otomatis untuk menentukan klaster terbaik dari akun instagram mahasiswa sebagai influencer penerimaan mahasiswa baru.

*Keywords: data mining; Instagram account; new student admission; influencer; the best cluster*

*Kata kunci: data mining; akun Instagram; penerimaan mahasiswa baru; influencer; klaster terbaik*

---

**Perancangan/metode/pendekatan:** Penelitian ini menerapkan metode web data extraction untuk mengekstraksi data akun Instagram mahasiswa dan K-Means untuk melakukan klusterisasi data, serta metode Silhouette Coefficient untuk menentukan jumlah kluster terbaik.

**Hasil:** Penelitian ini telah berhasil menentukan 7 akun mahasiswa terbaik dari 100 akun yang dapat dijadikan influencer penerimaan mahasiswa baru dengan nilai *Silhouette Score* yang tertinggi untuk jumlah influencer yang dipilih antara 5-10 yaitu sebesar 0.608 dari jumlah kluster sebanyak 22.

**Keaslian/ state of the art:** Penelitian terkait penerapan data mining untuk membangun sebuah model yang dapat melakukan ekstraksi data akun Instagram, serta dapat melakukan klusterisasi secara otomatis untuk menentukan kluster terbaik dari akun Instagram mahasiswa sebagai influencer penerimaan mahasiswa baru (PMB) ini belum pernah dilakukan pada penelitian-penelitian sebelumnya.

---

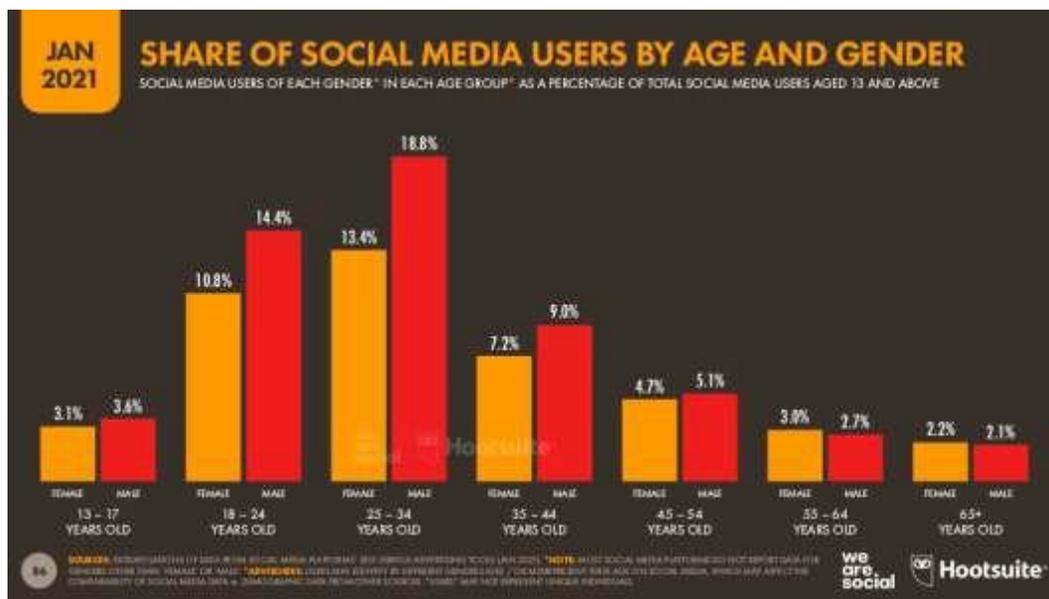
## 1. Pendahuluan

Perkembangan teknologi informasi saat ini telah banyak merubah kehidupan masyarakat, khususnya generasi muda atau yang sering disebut generasi milenial. Generasi ini memposisikan dunia digital sebagai area dalam mendapatkan dan membagikan segala jenis informasi [1]. Mereka cenderung sulit untuk dipisahkan dengan perangkat teknologi seperti *smartphone* [2]. Keberadaan jaringan internet dan kemudahan aksesnya semakin menambah tingginya konsumsi informasi dimana sebagian besar informasi mereka dapatkan dari media sosial [3]. Berdasarkan sebuah survei yang dilakukan oleh We Are Social dan Hootsuite, jumlah pengguna media sosial pada bulan Januari 2021 setara dengan lebih dari 53% dari jumlah total populasi dunia, dimana di Indonesia mencapai 65% [4]. Berdasarkan hasil survei tersebut, pengguna media sosial di dunia didominasi oleh pengguna dengan umur antara 18 sampai 34 tahun dimana secara detail ditunjukkan pada **Gambar 1**.

Tingginya tingkat penggunaan media sosial yang didominasi oleh generasi muda milenial ini tentunya memberikan peluang bagi suatu perguruan tinggi untuk dapat memanfaatkannya sebagai jalur promosi terkait penerimaan mahasiswa baru (PMB). Media sosial yang dapat diakses dari berbagai lokasi di Indonesia juga semakin mendukung penyampaian informasi kepada siswa SMA/SMK calon mahasiswa baru di seluruh Indonesia dimana rata-rata berumur kurang lebih 17 sampai 18 tahun. Adanya pandemi COVID-19 pada saat ini juga membuat proses promosi melalui media digital khususnya media sosial menjadi salah satu solusi yang efektif [5].

Instagram merupakan salah satu media sosial populer yang digunakan untuk berbagi gambar, yang tumbuh paling cepat dalam beberapa tahun terakhir [6]. Berdasarkan hasil survei yang diterbitkan oleh We Are Social dan Hootsuite pada bulan Januari 2021, Instagram menduduki peringkat ke lima media sosial paling banyak digunakan di dunia. Beberapa penelitian telah

mengeksplorasi kesuksesan Instagram sebagai alat bisnis, dimana dapat dimanfaatkan untuk mempromosikan produk dengan memanfaatkan seorang *influencer* [7], [8]. *Influencer* merupakan seorang yang memiliki peran sebagai pemberi pengaruh karena pendapat atau pandangannya terhadap suatu hal yang dapat memengaruhi keputusan para orang lain [9]. Pada umumnya, seorang dapat dikategorikan sebagai *influencer* pada sebuah media sosial jika seorang tersebut memiliki pengikut yang banyak atau memiliki *engagement rate* yang tinggi pada akun media sosialnya. *Engagement rate* dapat dihitung dengan mengukur persentase orang yang bereaksi terhadap terbitan pada akun media sosialnya dengan cara tertentu, seperti dengan menghitung jumlah "suka" atau "komentar" [10].



Gambar 1. Pengguna media sosial berdasarkan umur

Apabila dikaitkan dengan strategi pemasaran penerimaan mahasiswa baru melalui media sosial, pemanfaatan *influencer* sebagai media promosi melalui Instagram ini tentunya menjadi salah satu peluang yang dapat dimanfaatkan oleh suatu perguruan tinggi di era digital ini. Akan tetapi, pemanfaatan *influencer* yang telah memiliki reputasi tentunya akan memakan biaya yang tidak sedikit. Selain itu, diperlukan sebuah analisis untuk menentukan *influencer* yang sesuai untuk pemasaran penerimaan mahasiswa baru agar dapat lebih tepat sasaran. Sebagai salah satu solusi dari permasalahan tersebut, mendorong mahasiswa sebagai *influencer* menjadi salah satu langkah yang dapat diambil.

Permasalahan yang muncul selanjutnya adalah bahwa tidak semua mahasiswa dapat berperan sebagai *influencer* dan tidak semua akun Instagram dari mahasiswa dapat dimanfaatkan untuk media promosi. Hal ini dikarenakan tidak semua akun Instagram dari mahasiswa memiliki *engagement rate* yang cukup sebagai dasar penentuan *influencer*. Selain itu, isu terkait banyaknya pengikut palsu atau sering disebut *fake follower* dan juga pengikut pasif dari sebuah akun Instagram membuat pemilihan *influencer* menjadi lebih sulit. Dalam mengantisipasi hal tersebut, diperlukan analisis akun Instagram secara lebih dalam dimana dengan menentukan prioritas parameter penentuan *influencer*. Parameter seperti rata-rata jumlah suka, rata-rata jumlah komentar, dan tingkat keterbaruan terbitan dapat digunakan sebagai dasar tambahan

agar tidak hanya mengacu pada jumlah pengikut saja [11]. Parameter-parameter tersebut dapat dipilih sebagai dasar dalam mengklasterisasikan akun Instagram mahasiswa untuk menentukan kluster akun terbaik. Banyaknya mahasiswa yang memiliki akun Instagram tentunya akan membuat pengambilan nilai parameter dan proses klasterisasi tersebut akan memakan waktu yang lama jika dilakukan secara manual.

Berdasarkan permasalahan tersebut, maka dibutuhkan sebuah model yang dapat melakukan ekstraksi data akun Instagram, serta dapat melakukan klasterisasi secara otomatis agar tidak memakan waktu lama. Pada penelitian ini, ekstraksi data akun Instagram dapat dilakukan secara otomatis dengan memanfaatkan teknik seperti Web Scraping [12], sedangkan klasterisasi data akun Instagram berdasarkan parameter yang telah ditentukan dilakukan dengan metode klustering K-Means.

Penelitian terkait pemanfaatan media sosial seperti Instagram dalam mendukung proses promosi atau pemasaran penerimaan mahasiswa baru telah dilakukan dalam beberapa penelitian sebelumnya. Utami et al., di 2016 [13] melakukan analisis strategi komunikasi pemasaran seleksi mahasiswa baru Telkom University melalui Instagram. Dengan memanfaatkan terbitan promosi dalam bentuk foto yang dilengkapi dengan *caption* yang dapat membuat target audiens tidak bingung dan mengerti maksud dari foto yang dipromosikan, serta dilengkapi dengan *hashtag* yang dapat digunakan untuk mempermudah pencarian dalam mesin pencari Instagram.

Faradika et al., di 2020 [14] merancang sebuah sistem informasi penjadwalan otomatis pada Instagram untuk mendukung promosi program studi di Universitas Dharma Andalas. Pada penelitian tersebut disampaikan bahwa akun Instagram mampu menarik perhatian calon mahasiswa secara alami dengan cara menerbitkan konten yang tepat dan *up to date* untuk mendapatkan perhatian lebih. Pratama et al., di 2019 [15] melakukan analisis terkait pemanfaatan media sosial terhadap proses penerimaan mahasiswa baru di Institut Informatika dan Bisnis Darmajaya Bandar Lampung. Dalam penelitiannya yang menggunakan 100 mahasiswa baru sebagai sampel didapat bahwa terdapat 92.8% mahasiswa yang menggunakan Instagram. Dalam penelitian tersebut juga ditemukan bahwa pendekatan pemasaran melalui media sosial berdampak pada proses pemasaran penerimaan mahasiswa baru khususnya di era disrupsi.

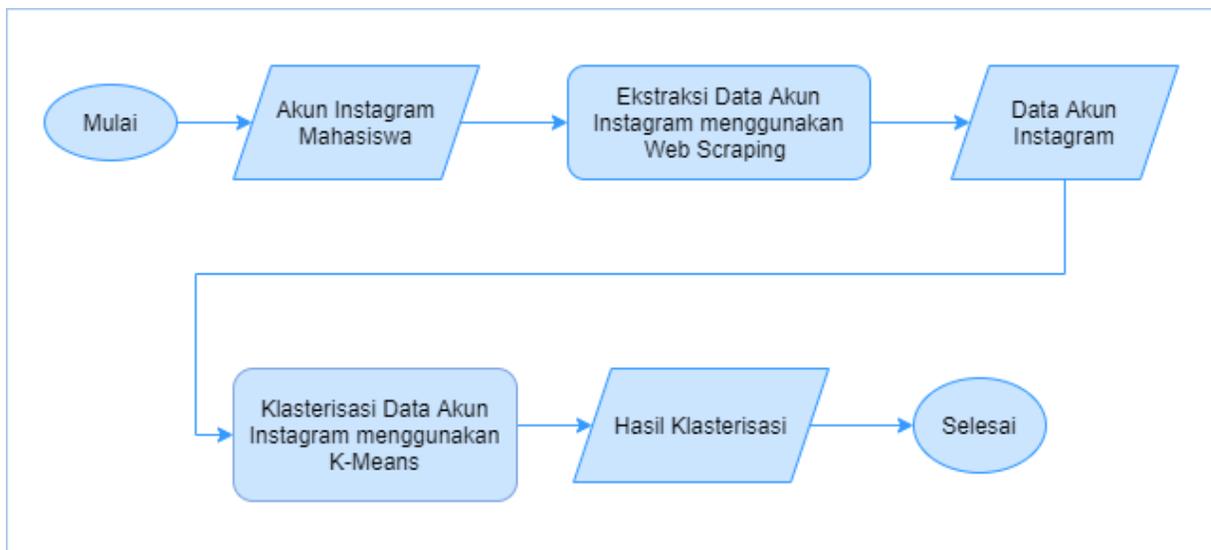
Penelitian terkait pemanfaatan teknologi data mining, khususnya terkait klasterisasi data dalam mendukung promosi penerimaan mahasiswa baru juga telah dimanfaatkan pada beberapa penelitian sebelumnya. Anggreini di 2019 [16] memanfaatkan metode K-Means untuk mengklasterisasi data calon mahasiswa yang mendaftar untuk menentukan strategi terbaik dalam melakukan promosi agar dapat lebih efektif dan efisien. Dengan memanfaatkan teknik klasterisasi, kegiatan promosi penerimaan mahasiswa baru dapat membuat penyebaran jumlah mahasiswa baru di tiap-tiap program studi menjadi merata. Kasri dan Jati di 2020 [17] memanfaatkan teknik klasterisasi dengan metode K-Means yang dikombinasikan dengan Simple Additive Weighting (SAW) untuk menentukan lokasi dan strategi pemasaran sebuah universitas. Dengan memanfaatkan kombinasi dari kedua metode tersebut, penentuan lokasi dan strategi pemasaran dapat dilakukan secara lebih optimal, khususnya dalam hal waktu, tenaga, dan biaya. Darma dan Nurcahyono di 2021 [18] memanfaatkan teknik klasterisasi dengan metode K-Medoids untuk menentukan teknik promosi dalam meningkatkan mutu kampus,

dalam rangka untuk mendapatkan mahasiswa baru. Dengan memanfaatkan metode tersebut, penentuan wilayah terbaik untuk promosi dapat dilakukan.

Penelitian terkait penerapan data mining untuk membangun sebuah model yang dapat melakukan ekstraksi data akun Instagram, serta dapat melakukan klusterisasi secara otomatis untuk menentukan kluster terbaik dari akun Instagram mahasiswa sebagai influencer penerimaan mahasiswa baru (PMB) ini belum pernah dilakukan pada penelitian-penelitian sebelumnya. Sehingga penelitian ini memberikan kebaruan yaitu terkait penerapan metode pada studi kasus yang baru.

## 2. Metode

Pada penelitian ini, metode penelitian yang digunakan adalah metode penelitian eksperimental. Metode penelitian tersebut digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendali [19]. Penelitian eksperimental pada penelitian ini ditujukan untuk mengetahui hasil penerapan metode Web Scraping dalam mengekstraksi data akun Instagram serta mengetahui hasil penerapan metode K-Means untuk melakukan klusterisasi data akun Instagram mahasiswa. Pada penelitian ini juga dilakukan studi pustaka sebagai sumber informasi baik secara cetak maupun elektronik untuk mendukung teori-teori yang digunakan sebagai dasar dalam penelitian. Adapun proses penerapan metode Web Scraping dan Data Mining untuk menentukan kluster terbaik dari akun Instagram mahasiswa sebagai influencer penerimaan mahasiswa baru dalam penelitian dapat dilihat pada **Gambar 2**.



**Gambar 2.** Alur proses model

### 2.1. Pengumpulan Data

Pada penelitian ini, pengumpulan data akun Instagram mahasiswa dilakukan dengan memanfaatkan Google Form. Dalam Google Form tersebut, terdapat kolom yang harus diisi oleh mahasiswa seperti kolom *username* Instagram dan status profil (*private* atau tidak). Data user Instagram mahasiswa yang terkumpul kemudian akan difilter berdasarkan status profil dimana akun yang status profilnya tidak *private* (*public*) yang akan digunakan. Hal ini

dikarenakan jika status profilnya *private*, maka informasi dari akun tersebut tidak dapat diekstraksi.

## 2.2. Metode Web Scraping

Metode Web Scraping merupakan sebuah metode yang dapat digunakan untuk mengekstrak data dari halaman web [20]. Pada penelitian ini, metode Web Scraping digunakan untuk mengekstrak data akun Instagram mahasiswa dari halaman web Instagram berdasarkan parameter yang telah ditentukan. Adapun parameter data yang akan di ekstrak adalah rata-rata jumlah suka, rata-rata jumlah komentar, tingkat keterbaruan terbitan, dan jumlah pengikut. Tahapan proses ekstraksi menggunakan metode Web Scraping terdiri dari analisis struktur data web dan pembuatan mesin *crawl* untuk mem-parsing dokumen HTML dan XML [21] dari halaman web. Penerapan metode Web Scraping pada penelitian ini dilakukan dengan memanfaatkan sebuah pustaka bernama Beautiful Soup yang tersedia dalam bahasa pemrograman Python.

## 2.3. Metode K-Means

Metode K-Means merupakan salah satu metode data mining yang digunakan untuk mengklusterisasi data. Metode K-Means merupakan sebuah metode analisis kluster yang mengarah pada pemartisian  $N$  objek pengamatan ke dalam  $K$  kluster (cluster) di mana setiap objek pengamatan dimiliki oleh sebuah kluster dengan nilai rata-rata atau mean terdekat [22]. Secara umum, metode K-Means ini dilakukan melalui beberapa tahapan seperti berikut [23]:

1. Penentuan jumlah kluster.
2. Penentuan pusat kluster (*centroid*) secara acak sesuai jumlah kluster yang telah ditentukan.
3. Penentuan kluster dari masing-masing data berdasarkan nilai kedekatan antara data dengan pusat kluster.
4. Hitung pusat kluster terbaru dari data yang ada di masing-masing kluster dengan cara mencari nilai rata-ratanya.
5. Penentuan kluster dari masing-masing data berdasarkan nilai kedekatan antara data dengan pusat kluster yang baru.
6. Ulangi langkah 4 sampai tidak ada data yang berpindah kluster.

Adapun perhitungan K-Means diperoleh dari **Persamaan 1** [24],

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (1)$$

dimana:

- $d$  : jarak
- $n$  : banyaknya objek
- $j$  : (dimulai dari 1 sampai  $n$ )
- $x_j$  : *feature* objek ke  $j$  terhadap  $x$
- $c_j$  : *centroid feature* ke  $j$

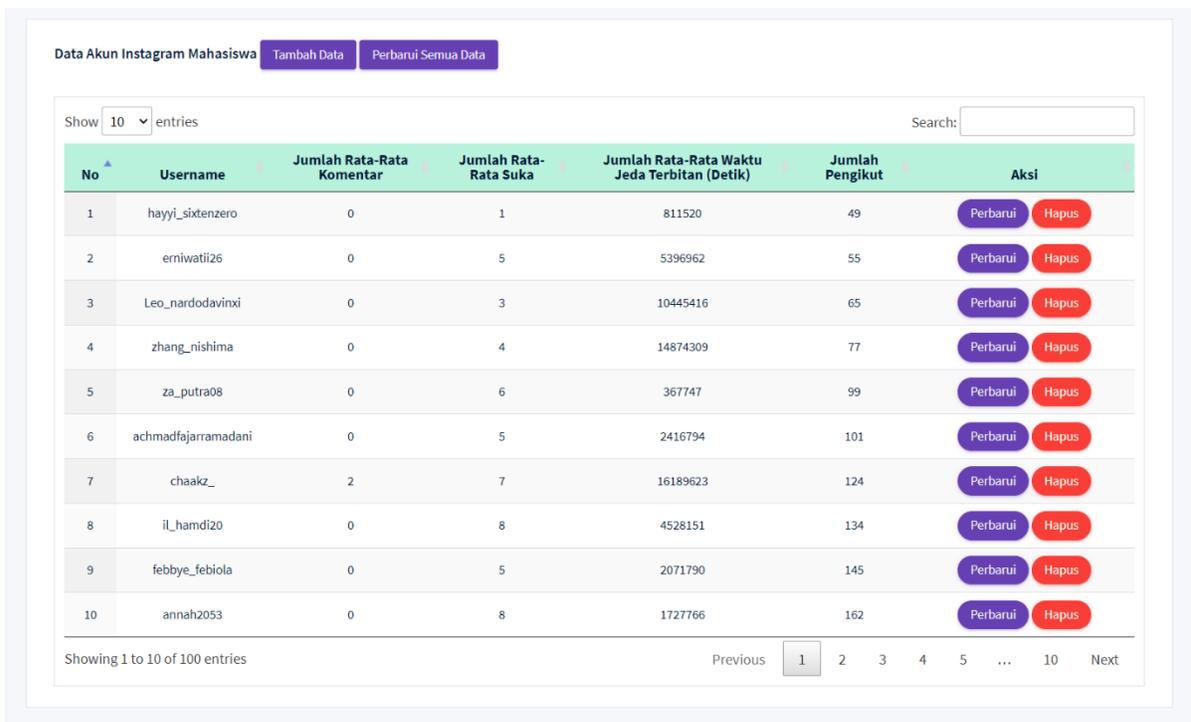
## 2.4. Evaluasi Hasil Ekstraksi dan Kualitas Kluster

Pada penelitian ini, evaluasi hasil ekstraksi dilakukan dengan mengukur sejauh mana metode Web Scraping diterapkan untuk mengekstrak data akun Instagram berdasarkan parameter yang telah ditentukan terhadap hasil yang sebenarnya. Data yang akan digunakan sebagai percobaan dalam pengujian ini terdiri dari data akun Instagram mahasiswa yang diperoleh dengan

memanfaatkan Google Form. Sedangkan evaluasi kualitas kluster dari hasil klusterisasi menggunakan metode K-Means dilakukan dengan metode Silhouette Coefficient. Metode ini dilakukan dengan cara rata-rata jarak antara sebuah objek dengan semua objek lain dalam kluster tersebut dan rata-rata jarak minimum dari sebuah objek ke semua kluster lain [23]. Jika nilai *Silhouette Score* dari 0 mendekati 1, maka hal ini berarti kluster yang berisi objek sangat padat dan objek terpisah jauh dari kluster-kluster lain dimana menunjukkan kualitas kluster semakin baik. Sebaliknya, jika *Silhouette Score* dari 0 mendekati -1, berarti kluster yang berisi objek tidak padat dan objek sangat dekat dengan kluster-kluster lain dimana menunjukkan kualitas kluster semakin tidak baik.

### 3. Hasil dan Pembahasan

Data nama akun Instagram mahasiswa yang digunakan pada penelitian ini diperoleh dengan memanfaatkan Google Form dimana terdapat 177 responden mahasiswa yang mengisi. Dari 177 nama akun tersebut, sebesar 56,5 % diantaranya merupakan akun yang berifat *public* yaitu sejumlah 100 akun. Oleh karena itu, data yang diproses pada penelitian ini adalah sebanyak 100 akun Instagram. Langkah selanjutnya adalah melakukan proses ekstraksi data dari web Instagram menggunakan teknik web scraping berdasarkan data nama akun Instagram tersebut untuk mengambil dan menghitung nilai rata-rata komentar, rata-rata suka, rata-rata waktu terbitan, dan jumlah pengikut. Hasil dari proses ekstaksi data pada penelitian ini ditunjukkan pada tampilan antar muka pada **Gambar 3**. Dari hasil evaluasi pada proses ekstraksi yang dilakukan dengan membandingkan hasil ekstraksi dengan metode Web Data Extraction terhadap hasil yang sebenarnya diperoleh bahwa metode tersebut dapat mengekstraksi nilai paramater yang telah ditentukan dari masing-masing data akun Instagram dengan benar.



No	Username	Jumlah Rata-Rata Komentar	Jumlah Rata-Rata Suka	Jumlah Rata-Rata Waktu Jeda Terbitan (Detik)	Jumlah Pengikut	Aksi
1	hayyi_sixtenzero	0	1	811520	49	Perbarui Hapus
2	erniwatiz6	0	5	5396962	55	Perbarui Hapus
3	Leo_nardodavinxi	0	3	10445416	65	Perbarui Hapus
4	zhang_nishima	0	4	14874309	77	Perbarui Hapus
5	za_putra08	0	6	367747	99	Perbarui Hapus
6	achmadfajarramadani	0	5	2416794	101	Perbarui Hapus
7	chaakz_	2	7	16189623	124	Perbarui Hapus
8	il_hamdiz0	0	8	4528151	134	Perbarui Hapus
9	febbye_febiola	0	5	2071790	145	Perbarui Hapus
10	annah2053	0	8	1727766	162	Perbarui Hapus

**Gambar 3.** Tampilan hasil ekstraksi data

Langkah selanjutnya adalah melakukan proses klusterisasi pada data akun Instagram tersebut berdasarkan parameter yang telah ditentukan yaitu nilai rata-rata komentar, rata-rata suka, rata-rata waktu terbitan, dan jumlah pengikut. Klusterisasi pada penelitian ini menerapkan metode K-Means dengan memanfaatkan Sklearn atau scikit-learn dimana merupakan alat sederhana dan efisien untuk analisis data. Sklearn adalah perpustakaan pembelajaran mesin perangkat lunak gratis untuk bahasa pemrograman Python. Pada metode K-Means ini dibutuhkan masukkan jumlah kluster yang akan dijadikan target. Pada metode ini, data akan secara otomatis dikelompokkan secara alami berdasarkan nilai rata-rata atau mean terdekat dari masing-masing data berdasarkan parameter yang telah ditentukan. Hasil dari proses ini akan menempatkan masing-masing data ke dalam sebuah label khusus sesuai kluster masing-masing. Hasil dari proses klusterisasi pada penelitian ini ditunjukkan pada **Gambar 4**.

No	Username	Jumlah Rata-Rata Komentar	Jumlah Rata-Rata Suka	Jumlah Rata-Rata Waktu Jeda Terbitan (Detik)	Jumlah Pengikut	Kluster
1	hayyi_sixtenzero	0	1	811520	49	21
2	erniwati26	0	5	5396962	55	3
3	Leo_nardodavinxi	0	3	10445416	65	4
4	zhang_nishima	0	4	14874309	77	16
5	za_putra08	0	6	367747	99	9
6	achmadfajarramadani	0	5	2416794	101	12
7	chaakz_	2	7	16189623	124	11
8	il_hamdi20	0	8	4528151	134	17
9	febbye_febiola	0	5	2071790	145	12
10	annah2053	0	8	1727766	162	0

**Gambar 4.** Hasil proses klusterisasi

Evaluasi kualitas kluster dari hasil klusterisasi pada penelitian ini menggunakan metode K-Means dilakukan dengan metode Silhouette Coefficient. Pada proses evaluasi ini, dilakukan percobaan dengan menginputkan beberapa nilai jumlah kluster. Adapun hasil dari percobaan tersebut dimana telah diurutkan berdasarkan nilai Silhouette dari yang tertinggi dapat dilihat pada Tabel 1. Pada proses ini juga akan ditentukan akun terpilih sebagai influencer dimana didasarkan pada kluster yang di dalamnya terdapat akun yang memiliki jumlah rata-rata komentar tertinggi. Hal ini ditentukan berdasarkan penelitian [11] dimana jumlah rata-rata komentar memiliki bobot tertinggi sebagai dasar penentuan akun Instagram untuk *endorsement*

atau dapat dikatakan *influencer* karena komentar merupakan respon yang dilakukan oleh pengikut terhadap postingan, yang dapat menjadi dasar interaksi kuat antara pengikut dan akun akun yang diikuti.

**Tabel 1.** Hasil percobaan klusterisasi dengan beberapa jumlah klaster

No.	<i>Silhouette Score</i>	Jumlah Klaster	Jumlah Influencer Terpilih
1.	0.6862530981	2	78
2.	0.6665132274	3	68
3.	0.6110001194	4	57
4.	0.6081819094	22	7
5.	0.6060393485	25	7
6.	0.6027770936	23	9
7.	0.6025611251	15	10
8.	0.6017935475	21	9
9.	0.6016921776	16	10
10.	0.6001461252	20	12
11.	0.5967271790	17	12
12.	0.5944964656	14	13
13.	0.5925944333	6	48
14.	0.5910859568	5	49
15.	0.5895918702	19	12
16.	0.5812122856	18	12
17.	0.5809599463	24	7
18.	0.5613173676	11	13
19.	0.5574623940	13	13
20.	0.5570567692	10	13
21.	0.5541503265	9	22
22.	0.5528497992	12	13
23.	0.5414573405	8	22
24.	0.5377279586	7	24

Berdasarkan hasil percobaan tersebut dapat dilihat jumlah klaster terbaik dari data tersebut adalah 2 dengan nilai *Silhouette Score* tertinggi yaitu 0.686. Akan tetapi jika kita lihat dari jumlah influencer terpilih, terdapat 78 dari 100 akun. Hal ini tentunya kembali kepada kebijakan perguruan tinggi akan memilih berapa mahasiswa yang dijadikan *influencer*. Sebagai contoh jika perguruan tinggi ingin memilih sekitar 5 sampai 10 mahasiswa yang akan dijadikan influencer, maka dapat dipilih jumlah klaster sebanyak 22 sehingga akan didapatkan 7 akun mahasiswa terbaik yang dapat dijadikan influencer penerimaan mahasiswa baru dengan nilai *Silhouette Score* yang tertinggi yaitu 0.608 untuk jumlah influencer yang dipilih antara 5-10. Hasil proses klusterisasi dengan jumlah klaster sebanyak 22 ditunjukkan pada **Gambar 5**.

The image shows a screenshot of a data analysis tool. The top section is titled "Data Kluster Akun Terbaik" and contains a table with 7 rows of data. The bottom section is titled "Silhouette Coefficient" and shows three key metrics: Silhouette Score (0.6081819094), Jumlah Kluster (22), and Jumlah Influencer Terpilih (7).

No	Username	Jumlah Rata-Rata Komentar	Jumlah Rata-Rata Suka	Jumlah Rata-Rata Waktu Jeda Terbitan (Detik)	Jumlah Pengikut	Kluster
1	anantayaputri	30	698	2989280	3558	19
2	aywlnsr	3	93	2484166	1675	19
3	candarr	2	134	2693812	1162	19
4	dells_nanik	1	15	2770413	237	19
5	yyaasss__	7	320	2705677	5552	19
6	_iamfadli	2	36	2753633	350	19
7	siti_sofiatur_m	0	32	2623500	284	19

Metric	Value
Silhouette Score	0.6081819094
Jumlah Kluster	22
Jumlah Influencer Terpilih	7

Gambar 5. Hasil proses klusterisasi dengan 22 kluster

#### 4. Kesimpulan dan Saran

Penelitian ini telah berhasil menerapkan metode Web Data Extraction untuk mengekstraksi data akun Instagram mahasiswa berdasarkan parameter yang telah ditentukan serta menerapkan metode data mining yaitu K-Means untuk melakukan klusterisasi data akun Instagram mahasiswa berdasarkan parameter yang telah ditentukan. Selanjutnya, penelitian ini juga telah berhasil menentukan 7 akun mahasiswa terbaik dari 100 akun yang dapat dijadikan influencer penerimaan mahasiswa baru dengan nilai *Silhouette Score* yang tertinggi untuk jumlah influencer yang dipilih antara 5-10 yaitu sebesar 0.608 dari jumlah kluster sebanyak 22. Sebagai kelanjutan dari penelitian ini dan untuk memperoleh hasil yang lebih baik lagi maka pada penelitian selanjutnya akan diterapkan beberapa metode data mining lainnya untuk mengklusterisasi data akun Instagram mahasiswa berdasarkan parameter yang telah ditentukan.

#### Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Politeknik Negeri Media Kreatif atas dukungan bantuan dana hibah penelitian internal dalam skema Penelitian Terapan Kompetitif 2021. Terima kasih dan apresiasi juga penulis sampaikan kepada pihak-pihak lain yang telah membantu sehingga penelitian ini dapat terselesaikan.

#### Daftar Pustaka

- [1] S. Sari, "LITERASI MEDIA PADA GENERASI MILENIAL DI ERA DIGITAL," *Prof. J. Komun. dan Adm. Publik*, vol. 6, no. 2, pp. 30–42, Dec. 2019, doi:

- 10.37676/professional.v6i2.943.
- [2] T. Mutia, “GENERASI MILENIAL, INSTAGRAM DAN DRAMATURGI: SUATU FENOMENA DALAM PENGELOLAAN KESAN,” *J. komunikasiana*, vol. 1, no. 1, pp. 38–47, Feb. 2018, doi: 10.24014/AN-NIDA.V41I2.4656.
- [3] A. Ambarwati and S. T. Raharjo, “Prinsip Kepemimpinan Character of A Leader pada Era Generasi Milenial,” *Philanthr. J. Psychol.*, vol. 2, no. 2, p. 114, Dec. 2018, doi: 10.26623/philanthropy.v2i2.1151.
- [4] W. A. Social and Hootsuite, “Digital 2021: The Latest Insights Into The ‘State of Digital’ - We Are Social,” 2021. [Online]. Available: <https://wearesocial.com/blog/2021/01/digital-2021-the-latest-insights-into-the-state-of-digital>. [Accessed: 18-Mar-2021].
- [5] P. Lestari and M. Saifuddin, “Implementasi Strategi Promosi Produk Dalam Proses Keputusan Pembelian Melalui Digital Marketing Saat Pandemi Covid’19,” *J. Manaj. dan Inov.*, vol. 3, no. 2, pp. 23–31, Aug. 2020, doi: 10.15642/manova.v3i2.301.
- [6] A. Priadana and M. Habibi, “Face detection using haar cascades to filter selfie face image on instagram,” in *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 2019, pp. 6–9, doi: 10.1109/ICAIT.2019.8834526.
- [7] A. Konstantopoulou, I. Rizomyliotis, K. Konstantoulaki, and R. Badahdah, “Improving SMEs’ competitiveness with the use of Instagram influencer advertising and eWOM,” *Int. J. Organ. Anal.*, vol. 27, no. 2, pp. 308–321, Apr. 2019, doi: 10.1108/IJOA-04-2018-1406.
- [8] S. Lee and E. Kim, “Influencer marketing on Instagram: How sponsorship disclosure, influencer credibility, and brand credibility impact the effectiveness of Instagram promotional post,” *J. Glob. Fash. Mark.*, vol. 11, no. 3, pp. 232–249, Jul. 2020, doi: 10.1080/20932685.2020.1752766.
- [9] M. R. Handika and G. S. Darma, “Strategi Pemasaran Bisnis Kuliner Menggunakan Influencer Melalui Media Sosial Instagram,” *J. Manaj. Bisnis*, vol. 15, no. 2, pp. 192–203, Apr. 2018, doi: 10.38043/JMB.V15I2.601.
- [10] R. Jaakonmäki, O. Müller, and J. vom Brocke, “The Impact of Content, Context, and Creator on User Engagement in Social Media Marketing,” in *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*, 2017, doi: 10.24251/hicss.2017.136.
- [11] M. I. Akrianto, A. D. Hartanto, and A. Priadana, “The Best Parameters to Select Instagram Account for Endorsement using Web Scraping,” in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2019, pp. 40–45, doi: 10.1109/ICITISEE48480.2019.9004038.
- [12] A. Himawan, A. Priadana, and A. Murdiyanto, “Implementation of Web Scraping to Build a Web-Based Instagram Account Data Downloader Application,” *IJID (International J. Informatics Dev.)*, vol. 9, no. 2, pp. 59–65, Oct. 2020, doi: 10.14421/IJID.2020.09201.

- 
- [13] M. A. Utami, M. T. Lestari, and B. P. S. Putri, "STRATEGI KOMUNIKASI PEMASARAN SMB TELKOM UNIVERSITY TAHUN 2015/2016 MELALUI MEDIA SOSIAL INSTAGRAM," *J. Sosioteknologi*, vol. 15, no. 2, pp. 309–318, Aug. 2016, doi: 10.5614/sostek.itbj.2016.15.02.13.
- [14] F. Faradika, R. Astri, and Z. Zulfahmi, "SISTEM INFORMASI PENJADWALAN OTOMATIS MEDIA SOSIAL INSTAGRAM UNTUK Mendukung Promosi Program Studi di Universitas Dharma Andalas," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 2, no. 2, pp. 225–230, Jul. 2020, doi: 10.47233/jteksis.v2i2.149.
- [15] Y. Aditiya Pratama, A. Wibasuri, L. Hakim, and I. nformatika dan Bisnis Darmajaya, "ANALYSIS OF THE SOCIAL MEDIA USAGE ON THE DECISION TO THE ADMISSION OF NEW STUDENT PROCESS IN PRIVATE UNIVERSITY (Case Study: IBI Darmajaya Bandar Lampung)," Dec. 2019.
- [16] N. L. Anggreini, "TEKNIK CLUSTERING DENGAN ALGORITMA K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, Dec. 2019, doi: 10.24036/tip.v12i2.215.
- [17] M. A. Kasri and H. Jati, "Combination of K-Means and Simple Additive Weighting in Deciding Locations and Strategies of University Marketing," *undefined*, 2020.
- [18] S. Darma and G. W. Nurcahyo, "Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids," *J. Inform. Ekon. Bisnis*, pp. 89–94, Mar. 2021, doi: 10.37034/infv3i3.87.
- [19] Sugiyono, *Metode Penelitian Pendidikan Pendekatan Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta, 2017.
- [20] R. C. Pereira and T. Vanitha, "Web Scraping of Social Networks," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 3, no. 7, pp. 237–240, 2015.
- [21] Fatmasari, Y. N. Kunang, and S. D. Purnamasari, "Web Scraping Techniques to Collect Weather Data in South Sumatera," in *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018*, 2019, doi: 10.1109/ICECOS.2018.8605202.
- [22] A. I. Abdullah, E. Winarko, and A. Musdholifah, "Metode Boost-K-means untuk Clustering Puskesmas berdasarkan Persentase Bayi yang Diimunisasi," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 4, no. 2, p. 89, Nov. 2020, doi: 10.30595/jrst.v4i2.7546.
- [23] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2017.
- [24] U. A. Nasron and M. Habibi, "Analysis of Marketplace Conversation Trends on Twitter Platform Using K-Means," *Compiler*, vol. 9, no. 1, pp. 51–62, May 2020, doi: 10.28989/compiler.v9i1.579.