

ANALISA TOPIK TERHADAP KOMENTAR MENGENAI METAVERSE MENGGUNAKAN METODE CLUSTERING K-MEANS

Andre Ertanto¹⁾, Viny Christanti Mawardi²⁾, Novario Jaya Perdana³⁾

¹⁾²⁾³⁾Teknik Informatika, Universitas Tarumanagara

Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

¹⁾andre.535190029@stu.untar.ac.id, ²⁾vinym@fti.untar.ac.id, ³⁾novariojp@fti.untar.ac.id

ABSTRACT

The development of the internet doesn't stop there, but continues to develop and evolve, even in a game, humans can interact with each other, make transactions with each other and maybe it becomes an opportunity to earn income, one that combines all of these things is known as the Metaverse. Metaverse is a layer that connects two worlds, namely: the real world and the virtual world. Metaverse offers a 3-dimensional experience that can be shared between users and interact within this technology where every activity of its users can be carried out with the help of Augmented and Virtual Reality technology services. In the metaverse, people want to see what topics are contained in the discussion. So a website was created to determine the topic of metaverse comments from social media. The method used on this website is Clustering K-means. Use this method to divide comments into groups that have something in common. The group of comments will be determined by the topic of the highest frequency of words. Evaluation uses the Elbow Method to determine the optimal k value in Clustering K-means.

Keyword: *Metaverse, Clustering, Elbow Method, Topic*

1. Pendahuluan

Teknologi informasi berjalan sangat pesat, hingga semuanya berada pada revolusi industri 4.0 dan *society* 5.0. Pada revolusi industri 4.0 merupakan fase teknologi yang mengubah cara beraktivitas manusia dalam skala, ruang lingkup, kompleksitas, dan transformasi dari pengalaman hidup. Sekarang masyarakat sudah berada pada *society* 5.0 dimana masyarakat berpusat pada manusia yang membuat seimbang antara kemajuan ekonomi dengan penyelesaian masalah sosial melalui sistem yang sangat menghubungkan melalui dunia maya dan dunia nyata.

Perkembangan digital industri juga berdampak terhadap perkembangan dari beberapa hal diantaranya perkembangan dari permainan, cara bertransaksi dan pola hidup dari masyarakat Indonesia pada khususnya

dan dunia pada umumnya. Perkembangan internet tidak berhenti sampai disitu saja, namun terus berkembang dan berevolusi, bahkan dalam sebuah game, manusia bisa saling berinteraksi, saling melakukan transaksi dan mungkin menjadi salah satu peluang untuk mendapatkan penghasilan, salah satu yang menggabungkan semua hal tersebut dikenal dengan nama Metaverse. Metaverse merupakan sebuah lapisan yang menghubungkan antara dua dunia, yaitu: dunia nyata dan dunia maya, Metaverse menawarkan sebuah pengalaman 3 dimensi yang dapat berbagi antar penggunaannya dan berinteraksi di dalam teknologi tersebut dimana setiap aktifitas dari penggunaannya dapat dilakukan dengan bantuan layanan teknologi *Augmented* dan *Virtual Reality*[2]. *Virtual Reality* adalah pemunculan gambar-gambar tiga dimensi yang dibuat komputer sehingga terlihat nyata dengan bantuan sejumlah peralatan tertentu, yang menjadikan penggunaannya seolah-olah terlibat langsung secara fisik dalam lingkungan tersebut[3]. *Augmented Reality* (AR) merupakan salah satu teknologi yang menggabungkan objek virtual dengan objek nyata ke dalam lingkup nyata tiga dimensi dan memproyeksikannya secara *real time*.

Metaverse merupakan salah satu hal yang sering dibicarakan di media sosial, dan bagaimana pengaruh dari implementasi *metaverse* tersebut pada kehidupan sehari-hari. Menjadi hal yang selalu ditunggu, untuk melihat bagaimana masyarakat menyikapi *metaverse*, tentunya dapat di analisa dengan melihat *opinion mining* yang muncul melalui media sosial, diantaranya Twitter, Instagram dan Youtube.

Text mining merupakan teknik yang digunakan untuk menggali data yang tersembunyi dari data yang berbentuk *text* [4]. *Opinion mining* juga merupakan salah satu teknik yang terdapat dalam *text mining*. Ada beberapa kategori yang termasuk didalam teknik *text mining* salah satunya adalah analisis sentimen atau *opinion mining*, yaitu suatu proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis, atau merupakan studi komputasi pendapat, perasaan dan emosi dalam bentuk teks. Ada beberapa algoritma atau metode yang digunakan untuk analisis

sentimen, antara lain Naïve Bayes, Support Vector Machine dan *clustering K-Means*.

Naïve Bayes memiliki kelebihan model yang cukup sederhana, tidak terlalu rumit dan bagus untuk mencari prediksi bersyarat. Kekurangan dari naïve bayes tersendiri adalah independen setiap atribut sehingga membuat akurasi berkurang dan nilainya harus 0. Support Vector Machine memiliki kelebihan akurasi yang lebih tinggi. Kelemahan metode ini sulit di implementasikan untuk himpunan dengan jumlah dimensi yang besar[6]. K-means juga memiliki kelebihan yaitu, cukup mudah dipahami dan diimplementasikan, proses pembelajaran membutuhkan waktu yang cepat dan sangat umum digunakan dalam teknik *clustering*. Di dalam penelitian ini algoritma yang digunakan adalah *K-Means* untuk mencari akurasi yang lebih baik dalam pengelompokan komentar dari setiap sosial media[7]. Dalam evaluasi juga menggunakan metode Elbow Method untuk mencari nilai k optimal pada *cluster*.

2. Metode Penelitian

Tahapan pembuatan sistem website Analisis Topik Komentar Metaverse dengan Clustering K-means adalah sebagai berikut:

2.1 Pengumpulan Data

Pengumpulan data ini akan diambil dari berbagai media sosial , yaitu Twitter, Youtube dan Instagram. Pengambilan data ini akan berjumlah 200 komentar dari masing-masing media sosial sehingga memiliki total 600 data menggunakan scrapping manual dan API. Data yang diambil memiliki topik Metaverse. Pengambilan data Youtube dan Twitter menggunakan API sedangkan Instagram scrapping manual sehingga datanya bersifat statis. Semua data yang didapatkan akan dimasukkan ke file excel. Data berisi komentar-komentar yang didapatkan dari media sosial.

2.2 Text Preprocessing

PreProcessing adalah tahap yang berguna untuk memproses sehingga data dapat lebih mudah dianalisa dan digunakan oleh komputer. Macam-macam langkah pre processing yang digunakan pada rancangan ini adalah Case Folding, Tokenizing, Stemming dan juga Filtering.

1. Case Folding

Case Folding adalah tahap *preprocessing* yang mengubah semua huruf dalam dokumen dari huruf yang besar menjadi huruf kecil yang bertujuan untuk semua

teks, hanya huruf ‘a’ sampai ‘z’. Karakter selain huruf tersebut dihilangkan dan dianggap sebagai pembatas.

2. Tokenizing

Tokenizing adalah tahap yang menjadikan sebuah kalimat menjadi kata dengan cara memecah kalimat. Setiap kata yang dipecah disebut juga dengan token. Proses ini melakukan penguraian deskripsi yang semula berupa kalimat berisi kata-kata dan tanda pemisah lain menjadi kata saja, baik merupakan kata penting maupun tidak penting.

3. Stopword Removal

Stopword removal adalah kata-kata dalam suatu kalimat yang sering muncul dalam jumlah banyak dan tidak memiliki arti, sehingga dengan dibuangnya kata-kata tersebut ukuran kosakata menjadi berkurang dan harapannya sisa dari kata-kata memiliki bobot yang tinggi. Contoh kamus *stopword removal* dapat dilihat di Gambar 1.



Gambar 1 *Stopword Removal*

4. Stemming

Stemming adalah tahap mengembalikan kata menjadi kata dasar yang telah dihasilkan dari tahap *stopword removal*. Pada tahap ini kata akan dilakukan pemotongan kata imbuhan. *Stemming* pada bahasa Indonesia akan lebih kompleks dibandingkan dengan bahasa Inggris dikarenakan tidak ada rumus pasti¹.

2.3 TF-IDF

Dalam preprocessing juga terdapat tahap TF-IDF yang digunakan untuk menghitung bobot setiap kata yang paling umum pada *information retrieval*. Metode ini menggabungkan dua konsep perhitungan, yaitu frekuensi data dalam dokumen tertentu dan frekuensi data *inverse* dari dokumen yang mengandung kata tersebut. TF adalah frekuensi di mana sebuah kata muncul dalam dokumen. Sedangkan IDF adalah sebuah kata yang jarang muncul atau kebalikan dari TF dalam dokumen. Perhitungan TF-IDF termasuk sangat efisien, mudah dan memiliki hasil yang akurat. Metode tf-idf dirumuskan sebagai berikut[8]:

$$w_{dt} = t_{fdt} * i_{dft} \quad (1)$$

¹ Ibid.

$$\text{idf} = \log N \frac{N}{\text{dft}} \quad (2)$$

Keterangan:

wdt= bobot dokumen ked terhadap kata keterangan
 tfdt = banyaknya kata yang dicari pada sebuah dokumen
 idft = Inversed Document Frequency ($\log(N/\text{df})$)
 N = total dokumen
 dft= banyak dokumen yang mengandung kata yang dicari.

2.4 Clustering K-Means

Metode k-means berawal dari penentuan jumlah kelompok yang ingin dibentuk, kemudian menentukan objek sebagai centroid awal yang biasanya dilakukan secara acak, selanjutnya menghitung ukuran jarak dari masing-masing data ke centroid. Setelah data masuk pada centroid terdekat dan membentuk kelompok baru, centroid baru ditentukan kembali dengan menghitung rata-rata objek pada centroid yang sama. Jika masih ada perbedaan dengan centroid yang sudah dibentuk, maka dilakukan perhitungan kembali centroid baru. Hasil cluster dengan metode k-means sangat bergantung pada nilai pusat kelompok awal yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan kelompok yang berbeda. Ada beberapa cara memberi nilai awal misalnya dengan mengambil sampel awal dari objek, lalu mencari nilai pusatnya, memberi nilai awal secara random, menentukan nilai awalnya atau menggunakan hasil dari kelompok hierarki dengan jumlah kelompok yang sesuai. Langkah-langkah algoritma *clustering k-means* adalah sebagai berikut:

1. Inisialisasi: tentukan nilai k sebagai jumlah klaster yang diinginkan dan matriks jarak yang diinginkan.
2. Pilih k data dari set data X sebagai centroid.
3. Alokasikan semua data ke centroid terdekat dengan matriks jarak yang sudah ditetapkan
4. Hitung kembali centroid berdasarkan data yang mengikuti klaster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu tidak ada data yang berpindah klaster.

Tujuan dari algoritma *k-means* adalah meminimumkan jarak antara data dengan centroid yang terdekat. Berikut merupakan rumus dari *clustering k-means dengan Euclidean distance*. Berikut Rumus Euclidean Distance[9]:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (3)$$

Keterangan:

d_{ij} = Jarak data i ke pusat *cluster* j
 x_{ik} = nilai objek i pada variabel ke-k
 x_{jk} = nilai objek j pada variabel ke-k
 n = banyaknya variabel yang diamati

2.5 Elbow Method

Metode Elbow merupakan salah satu metode yang digunakan untuk menentukan cluster yang optimal melalui perbandingan hasil presentase antara jumlah cluster yang akan membentuk siku pada grafik. Jika perbandingan nilai cluster pertama dengan nilai cluster kedua menghasilkan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai cluster tersebut yang tepat.

Untuk mendapatkan perbandingannya adalah dengan menghitung Sum of Squared Error(SSE) dari masing -masing nilai cluster. Karena semakin besar jumlah nilai cluster k, maka nilai SSE akan semakin kecil. Berikut rumus SSE[10]:

$$SSE = \sum_{K=1}^K \sum |x_i - c_k|^2 \quad (4)$$

Keterangan:

K = cluster ke-c
 X_i = jarak data ke obyek-i
 C_k = pusat cluster ke-i

3. Hasil Pengujian

3.1. Tampilan

1. Modul Beranda

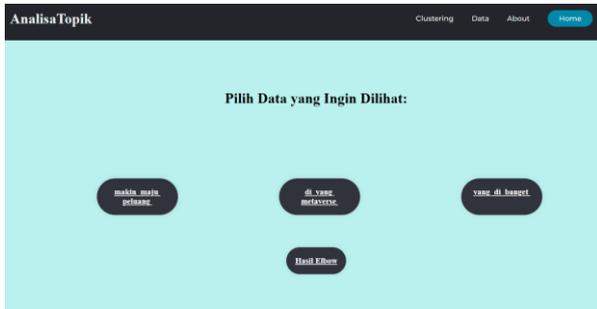
Modul beranda merupakan halaman pertama yang pertama kali dilihat saat memasuki website Analisa Topik. Dalam halaman tersebut terdapat tombol navbar yang dapat mengalihkan halaman beranda ke halaman lain dan juga terdapat tombol Twitter, Youtube dan Instagram yang mengalihkan halaman ke halaman clustering. Modul beranda dapat dilihat pada **Gambar 2**.



Gambar 2 Modul Beranda

2. Modul Clustering

Modul clustering merupakan halaman yang berisi tombol topik dari hasil clustering yang sudah dikelompokkan. Terdapat juga tombol hasil elbow method. Modul clustering dapat dilihat pada **Gambar 3**.



Gambar 3 Modul Clustering

3. Modul Komentar

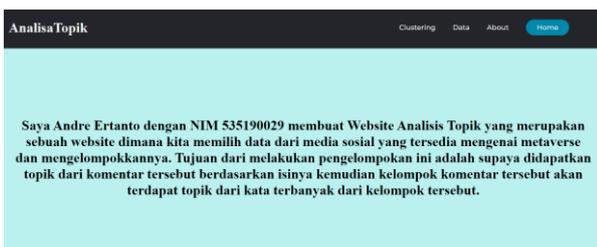
Pada modul komentar ini terdapat komentar yang sudah dikelompokkan mengenai metaverse. Modul komentar dapat dilihat pada **Gambar 4**.



Gambar 4 Modul Komentar

4. Modul About

Pada modul about ini berisi informasi mengenai pembuat *website*, deskripsi dan juga manfaat dari *website*. Modul about dapat dilihat pada **Gambar 5**.



Gambar 5 Modul About

3.2. Pengujian Sistem

Data yang telah diambil akan di preprocessing lalu akan TF-IDF. Setelah diproses akan di clustering dalam penentuan k optimal pada clustering menggunakan metode Elbow Method.

Tabel 1 Hasil Elbow Instagram

K	Instagram	
	Distortions	Inertia
1	50.027	666843.745
2	25.048	166841.893
3	16.733	74258.724
4	12.583	41862.976
5	10.101	26876.969
6	8.454	18764.131
7	7.273	13827.246
8	6.398	10656.955

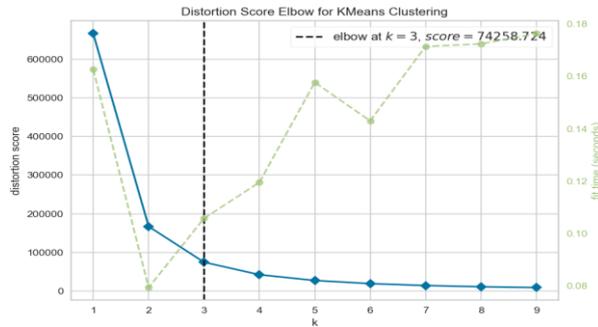
Tabel 2 Hasil Elbow Twitter

K	Twitter	
	distortions	Inertia
1	50.0275	666842.460
2	25.0484	166841.190
3	16.733	74258.081
4	12.591	41938.560
5	10.103	26897.571
6	8.451	18749.237
7	7.281	13870.348
8	6.398	10659.042

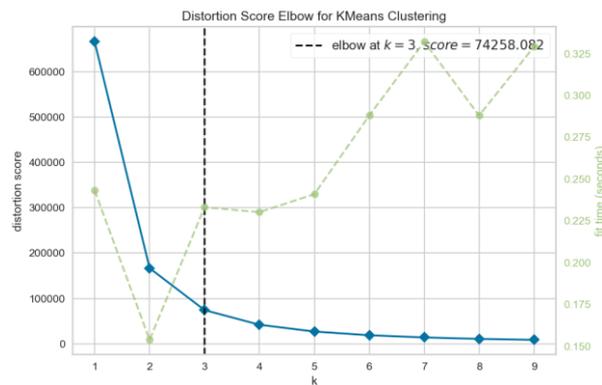
Tabel 3 Hasil Elbow Youtube

K	Youtube	
	distortions	Inertia
1	49.778	656892.404
2	24.924	164365.130
3	16.653	73186.219
4	12.528	41300.791
5	10.057	26507.573
6	8.412	18469.465
7	7.237	13612.784
8	6.368	10497.680

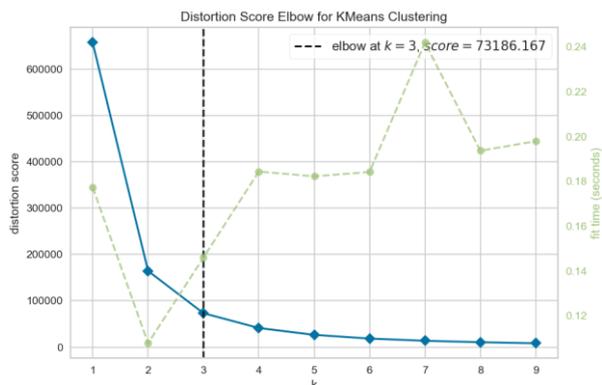
Dapat dilihat pada **Tabel 1**, **Tabel 2**, dan **Tabel 3** merupakan nilai dari elbow distortion dan inertia dari k1-k8 dari masing-masing media sosial. *Distortion* adalah rata-rata jarak ke *eulclidean distance* dari pusat *centroid cluster* masing-masing. Sedangkan *inertia* adalah hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat *cluster*.



Gambar 7 Grafik Elbow Instagram



Gambar 6 Grafik Elbow Twitter



Gambar 8 Grafik Elbow Youtube

Dapat dilihat pada **Gambar 6**, **Gambar 7** dan **Gambar 8** merupakan grafik elbow pada setiap media sosial. Pada grafik diatas dapat disimpulkan bahwa nilai optimal k untuk masing-masing media sosial adalah 3. Grafik juga memudahkan dalam pencarian k terbaik yang terjadi penurunan drastis, stabil dan membentuk sudut siku.

4. Kesimpulan

Berdasarkan hasil pembuatan website untuk Analisa topik dengan metode Clustering K-means, diperoleh sebagai berikut:

1. Aplikasi yang dibuat untuk mengelompokkan komentar metaverse menjadi beberapa topik

dengan menggunakan metode Clustering K-means dan Elbow Method.

2. Data yang digunakan memiliki total 600 komentar metaverse dari setiap media sosial Twitter, Youtube dan Instagram yang masing-masing berjumlah 200 komentar. Data diambil dengan cara scrapping dan API.
3. Penentuan topik dari setiap cluster didapatkan dengan cara mencari frekuensi kata terbanyak dan ditampilkan hanya 3 kata.
4. Topik yang didapatkan dapat digunakan untuk penelitian yang lain ataupun digunakan untuk *googling*.
5. Dalam penentuan nilai k yang optimal menggunakan Elbow Method. Masing-masing media sosial memiliki nilai k optimal yaitu, 3. Sehingga tidak ada perbedaan dari nilai k dari setiap media sosial.

REFERENSI

- [1] Puspita, Yenny; Fitriani, Yessi; Astuti, Sri dan Novianti, Sri, 2020, Selamat Tinggal Revolusi Industri 4.0, Selamat Datang Revolusi Industri 5.0, <https://jurnal.univpgri-palembang.ac.id/index.php/Prosidingpps/article/view/3794/3565>, tanggal akses 29 September 2022.
- [2] Damar, Muhammet, 2021, Metaverse Shape of Your Life for Future: A Bibliometric Snapshot. <http://arxiv.org/abs/2112.12068>, diakses tanggal 10 September 2022.
- [3] Erwanto; Bagus; Salsabil, Aditya Brian; Muzayyana, Nurul dan Hartanti, Dwi, 2022, Virtual Reality : Analisis Minat Teknologi Interaktif Kunjungan Wisatawan. <http://ojs.udb.ac.id/index.php/Senatib/article/download/1953/1538>, diakses tanggal 16 September 2022.
- [4] Zahrotun, Lisna, 2017, Perancangan Text Mining Pengelompokan Penelitian Dosen Menggunakan Metode Shared Nearest Neighbor Dengan Euclidean Similarity, <http://jurnal.umk.ac.id/index.php/SNA/article/view/1459>, diakses tanggal 20 September 2022.
- [5] Li, Gang, dan Liu, Fei, 2010, A Clustering-Based Approach on Sentiment Analysis, <https://doi.org/10.1109/ISKE.2010.5680859>, diakses tanggal 10 September 2022.
- [6] Muhammad Ichwan, Irma Amelia Dewi dan Zeni Muharom S, 2019, Klasifikasi Support Vector Machine (SVM) Untuk Menentukan Tingkat Kemanisan Mangga Berdasarkan Fitur Warna, *MIND Journal*, No. 2, Vol.3, hal 16-23.
- [7] Faesal; Andris; Muslim, Aziz; Ruger, Aditya Hastami dan Kusri, 2020, Sentimen Analisis Terhadap Komentar Konsumen Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means, *Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer (MATRIK)*, No.2, Vol.19, hal 207-213.
- [8] Westley, Vincentius, Dimitrius Thomas, and Fitrah Rumaisa, 2022, Analisis Sentimen Ulasan Hotel Bahasa

Indonesia Menggunakan Support Vector Machine Dan TF-IDF, *Jurnal Media Informatika Budidarma*, No. 3, Vol. 6, hal 16-23.

- [9] R Suwanda, Zulfahmi Syahputra dan Elvi M Zamzami, 2020, Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K, <https://iopscience.iop.org/article/10.1088/1742-6596/1566/1/012058/meta>, diakses tanggal 24 Oktober 2022
- [10] Prima Gabriel Ryan, 2021, Analisa Perbandingan Nilai K Terbaik Untuk Clustering K-MEANS Menggunakan Pendekatan Elbow dan Silhoutte Pada Citra Aksara Jawa, https://repository.usd.ac.id/40190/2/175314084_full.pdf, diakses akses 24 Desember 2022

Andre Ertanto, mahasiswa di Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Viny Christanti M., M.Kom, memperoleh gelar S.Kom melalui penempuhan ilmu di Universitas Tarumanagara, Jakarta tahun 2004 dan gelar M.Kom melalui penempuhan ilmu di Universitas Indonesia, Depok tahun 2008. Saat ini bekerja sebagai Staf Pengajar Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.

Novario Jaya Perdana, memperoleh gelar S.Kom melalui penempuhan ilmu di ITS tahun 2011 dan M.T. melalui penempuhan ilmu di Universitas Indonesia tahun 2016. Saat ini bekerja sebagai Staf Pengajar Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara, Jakarta.