



**PENERAPAN KLASTERISASI DENGAN MENGGUNAKAN ALGORITMA K-MEANS
UNTUK MENGETAHUI TINGKAT KUNJUNGAN WISATA DI BOJONEGORO**

Oleh

Devi Dwi Nuraliza¹⁾, Ita Aristia Sa'ida²⁾

^{1,2}Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro

Jl. Jenderal Ahmad Yani no. 10 Bojonegoro

Email: ¹devidwinuraliza@gmail.com, ²itaaristia@unugiri.ac.id

Abstrak

Bojonegoro is one of the regencies in East Java that has a tourist attraction that is in demand by tourists, tourist visits in Bojonegoro can increase the country's foreign exchange and improve the people's economy. The addition of tourism potential can affect the number of visits which leads to a reduction in visits to several tourist sites. However, the Covid-19 pandemic has had a fairly serious impact on the Indonesian economy, especially in the tourism sector so that the growth in the rate of tourist visits in Bojonegoro from the amount of data every month and year is still very difficult to know. Clustering of tourist visits needs to be done to find out the clusters of the highest level of object visits to the lowest visits so that improvements can be made to facilities and infrastructure of superior objects that can increase the number of visits and have an impact on increasing State Foreign Exchange. The clustering method uses the K-Means algorithm with distance calculations using Euclidean distance. The data source comes from the Department of Culture and Tourism of Bojonegoro Regency. The research data used is the number of tourist visits in 2016-2020. The data are grouped into 3 clusters of tourist visits, namely high, medium and low. The results of the Euclidean Distance calculation are three (3) iterations with a fixed cluster of 16 and a cluster changing of 0. Clustering in 2016, 2017 and 2018 the highest tourist visit value (C1) is 1. Moderate tourist visit (C2) is 1 and tourist visit low (C3) as many as 14. The results of the Euclidean Distance calculation in 2019 obtained the highest visit value (C1) as much as 1, moderate tourist visits (C2) as many as 3 and low tourist visits (C3) as many as 12. While the results of the Euclidean Distance calculation in 2020 obtained The highest tourist visit value (C1) is 3. Moderate tourist visits (C2) are 2. While low tourist visits (C3) are 11.

Keywords : Tourist Visits, Clustering, K-Means Algorithm

PENDAHULUAN

Dinas Kebudayaan dan Pariwisata (DISBUDPAR) wilayah Kabupaten Bojonegoro, Jawa Timur memiliki tugas untuk melaksanakan urusan pemerintahan Kabupaten Bojonegoro dalam bidang budaya dan pariwisata berdasarkan asas otonomi daerahnya. Adapun tugas Disbudpar adalah sebagai pelaksana urusan pemerintah daerah pada bidang pariwisata dan pelestarian budaya di wilayah kerjanya. Fungsi Disbudpar ialah merumuskan kebijakan bidang pariwisata, kesenian, kebudayaan dan perfilman,

penyelenggara pariwisata dan kebudayaan, pembinaan dan pembimbingan pada pelaku pariwisata dan budaya di wilayah kerjanya, koordinator UPTD, hingga pelaporan dan koordinasi urusan pariwisata dan budaya [1]. Kunjungan wisatawan di Bojonegoro dapat meningkatkan devisa negara dan meningkatkan perekonomian masyarakat di Daerah Wisata. Bojonegoro merupakan salah satu Kabupaten di Jawa Timur yang memiliki daya tarik wisata yang diminati oleh wisatawan, dilihat dari jumlah kunjungan wisatawan yang semakin naik dari tahun ke tahun. Berbagai macam



objek wisata yang ditawarkan di Kabupaten Bojonegoro, mulai dari wisata alam, wisata budaya, wisata religi dan wisata kuliner. Penambahan potensi wisata dapat mempengaruhi jumlah kunjungan yang mengarah pada pengurangan kunjungan di beberapa lokasi wisata. Hal ini karena fokus wisatawan akan terpecah ke beberapa potensi wisata baru dimana dapat dibuktikan dari data peningkatan potensi wisata dan jumlah kunjungan dari tahun ke tahun. Akan tetapi Pandemi Covid-19 membawa dampak yang cukup serius bagi perekonomian bangsa Indonesia, terutama pada sektor pariwisata sehingga pertumbuhan tingkat kunjungan wisatawan di Kabupaten Bojonegoro dari banyaknya data kunjungan wisatawan pada setiap bulan dan tahunnya masih sangat sulit untuk diketahui dan berdasarkan wawancara yang telah dilakukan dengan pihak Dinas Kebudayaan dan Pariwisata Bojonegoro bahwa belum terdapat sistem untuk mengetahui banyaknya kunjungan sebagai acuan pengembangan wisata kedepannya.

Dari pemikiran di atas maka, dilakukan penelitian tentang *data mining* pada kunjungan wisatawan di Kabupaten Bojonegoro. *Data Mining* adalah analisa yang dilakukan secara otomatis pada data besar dan kompleks dengan tujuan untuk mendapatkan pola penting yang keberadaannya biasanya tidak disadari [2]. *Data mining* merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi [3]. Tujuannya adalah untuk mengetahui objek wisata yang memiliki potensi yang paling rendah dalam kunjungan wisatawan sampai yang paling tinggi dimana penelitian ini juga dapat mengetahui kuantitas atau jumlah pengunjung di tempat wisata dan golongan-golongan pecinta wisata. Dalam data mining juga terdapat metode-metode yang dapat digunakan salah satunya yaitu *clustering*. *Clustering* atau klasterisasi adalah metode

pengelompokan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [4].

Salah satu algoritma *clustering* adalah *K-Means* yang merupakan salah satu metode dalam fungsi *clustering* atau pengelompokan. Menurut (Larose, 2005) *clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar kemiripan objek yang diteliti dimana *K-Means* merupakan suatu algoritma yang digunakan dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda – beda. Algoritma *K-Means* juga sangat mudah di implementasikan serta menggunakan prinsip yang sederhana dan dapat di jelaskan dalam *non-statistik* dan Algoritma *K-Means* ini sangat cocok untuk penelitian yang bersifat *unsupervised* (tanpa arahan). Dalam hal ini, penerapan Data mining mampu menjadi solusi dalam menganalisa data.

Dengan data yang sudah dikelompokkan menggunakan Algoritma *K-Means* diharapkan dapat mempermudah Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro dalam klasterisasi hasil jumlah kunjungan wisatawan dan dapat menjadi acuan peningkatan kunjungan wisatawan di Bojonegoro pada lokasi wisata yang sepi kunjungan. Keberhasilan pembangunan Pariwisata ditandai dengan meningkatnya arus kunjungan wisatawan yang akan memberikan manfaat kesejahteraan bagi masyarakat luas, bagi usaha Pariwisata dan usaha terkait serta Pemerintah selaku pembina Pariwisata di Daerah sehingga dapat dilakukan perbaikan sarana dan prasarana objek wisatawan unggulan yang dapat meningkatkan jumlah kunjungan wisatawan yang berdampak pada pengenalan objek wisata dan peningkatan *devisa* Negara.



LANDASAN TEORI

1. Kunjungan Wisata di Kabupaten Bojonegoro

Kunjungan wisatawan di Bojonegoro dapat meningkatkan devisa negara dan meningkatkan perekonomian masyarakat di Daerah Wisata. Keberhasilan pembangunan Pariwisata ditandai dengan meningkatnya arus kunjungan Wisatawan yang akan memberikan manfaat kesejahteraan bagi masyarakat luas, bagi usaha Pariwisata dan usaha terkait serta Pemerintah selaku Pembina Pariwisata di Daerah sehingga dapat dilakukan perbaikan sarana dan prasarana objek wisatawan unggulan yang dapat meningkatkan jumlah kunjungan wisatawan yang berdampak pada pengenalan objek wisata dan peningkatan *devisa* Negara.

2. Data Mining

Data Mining adalah analisa yang dilakukan secara otomatis pada data besar dan *kompleks* dengan tujuan untuk mendapatkan pola penting yang keberadaannya biasanya tidak disadari [2]. *Data Mining* melakukan *ekstraksi* untuk mendapatkan informasi penting yang sifatnya *implisit* dan sebelumnya tidak diketahui, dari suatu data (Witten et al., 2011). Sedangkan menurut Berry, pengertian *data mining* merupakan aktivitas menganalisis data dalam jumlah yang besar untuk menemukan *pattern* (pola) dan *rule* (aturan) yang berguna.

Data mining merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu *data mining* memiliki peranan yang sangat penting dalam beberapa bidang kehidupan, diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi [5] untuk mengekstraksi dan mengidentifikasi informasi dalam pengolahan data berskala besar, dimensi yang tinggi, dan data yang berbeda sifat. Dalam *data mining* terdapat metode-metode yang dapat digunakan seperti *Classification*, *Clustering*, *Association*, *Regression*, *Forecasting*, *Sequence Analysis*, dan *Deviation Analysis*.

3. Unsupervise Learning

Dalam pembelajaran *machine learning*, terdapat skenario-skenario seperti *Unsupervised learning*. *Unsupervised learning* merupakan Algoritma *data mining* mencari pola dari semua *variable (atribut)*. *Variable (atribut)* yang menjadi *target/label/class* tidak ditentukan (tidak ada). Algoritma *clustering* adalah algoritma *unsupervised learning*. [6]

Unsupervise Learning merupakan jenis data tidak terarah dimana data tidak memiliki label dan hanya mempunyai *variabel input* tapi tidak mempunyai *variabel output* yang berhubungan. Metode ini hanya akan mempelajari suatu data berdasarkan kedekatannya saja atau yang biasa disebut dengan *clustering*. Metode *unsupervised learning* yang paling umum adalah analisis *cluster*, yang digunakan pada analisa data untuk mencari pola-pola tersembunyi atau pengelompokan dalam data. Salah satu contoh *implementasi unsupervised learning* adalah *clustering* dengan algoritma yang dapat digunakan dalam *unsupervised learning* seperti *K-Means*.

4. Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk mengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. *Clustering* adalah metode yang digunakan dalam *data mining* yang cara kerjanya mencari dan mengelompokan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya. *Clustering* mempartisi *dataset* kedalam *subset*, analisis *cluster* data yang terdapat dalam satu *cluster* memiliki tingkat kesamaan yang tinggi dan data yang terdapat dalam satu *cluster* yang berbeda memiliki tingkat kesamaan yang rendah. Ciri khas dari teknik *data mining* ini adalah mempunyai sifat tanpa arahan (*unsupervised*), yang dimaksud adalah teknik ini diterapkankan



tanpa perlunya data training dan tanpa ada teacher serta tidak memerlukan target *output*.

5. Algoritma K-Means

K-Means merupakan salah satu metode pengelompokan data *non-hirarki* (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok [7]

Menurut MacQueen, *K-Means* merupakan metode klasterisasi yang paling terkenal dan banyak digunakan dalam berbagai bidang karena bentuknya yang sangat sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengklaster data yang cukup besar [8]. *K-Means clustering* merupakan sebuah metode dari *unsupervised learning*. Metode perhitungan jarak pada *K-Means* meliputi *Eclidean Distance*, *Manhattan Distance* dan *Minkowski Distance* dimana perhitungan jarak terdekat data dengan titik pusat data/*centroid* yaitu dengan rumus *distance space*.

Pengelompokan data dengan langkah-langkah pengukuran jarak *Euclidean Distance* pada *K-Means* :

- a) Melakukan normalisasi data untuk memetakan nilai dalam rentang 0-1.
- b) Menentukan jumlah *cluster* yang ingin dibentuk
- c) Alokasikan data ke dalam *cluster* (penentuan centroid awal) secara random atau bisa menggunakan diagram.
- d) Hitung *centroid*/ rata-rata terdekat dari data yang ada di masing-masing *cluster* dengan persamaan *Euclidean Distance* atau metode pengukuran jarak lainnya :

$$e) D_{L1}(X_2, X_1) = |X_2 - X_1| = \sqrt{\sum_{j=1}^p (X_{2j} - X_{1j})^2}$$

- f) Kelompokan masing-masing data ke dalam cluster dengan centroid/rata-rata jarak paling pendek.
 - g) Kembali ke Step 3, apabila masih ada data yang berpindah cluster sampai sudah tidak ada lagi data yang berpindah ke cluster yang lain.
6. Pre-processing

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang *inkosisten*, dan memperbaiki kesalahan pada data. *Preprocessing* merupakan tahapan persiapan data yang tidak terstruktur menjadi terstruktur yang akan dilakukan ke tahapan lainnya. Tahapan data *pre-processing* adalah dengan metode *Missing value* atau dapat diartikan sebagai data atau informasi yang “hilang” atau tidak tersedia, *inconsistent* data atau duplikasi data yang menyebabkan data menjadi tidak konsisten, dan *noisy* data atau data *outlier* yang disebabkan oleh user yang melakukan kesalahan saat melakukan *input*. *Preprocessing* data dilakukan dengan data *cleaning* yang masuk pada 3 kategori tersebut akan dihapus.

METODE PENELITIAN

Metodologi penelitian dilakukan dengan Langkah-langkah sebagai berikut:

- a. Studi Literatur
Pencarian literatur teori yang relevan dengan kasus atau permasalahan yang ditemukan dalam rangka melengkapi pengetahuan dasar dan teori-teori melalui buku-buku, jurnal yang berkaitan dengan masalah dan tujuan penelitian. Output Study Literatur adalah terkoleksinya referensi yang relevan dengan perumusan masalah.
- b. Pengumpulan Data
Pengumpulan data melalui penelitian dan wawancara yang dilakukan pada Kantor

<http://ejurnal.binawakya.or.id/index.php/MBI>



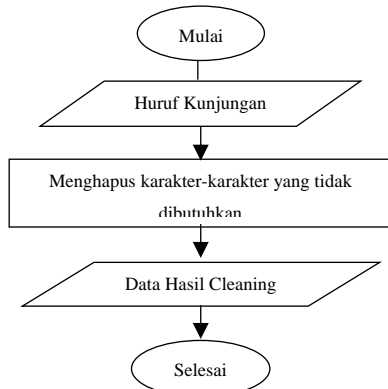
Dinas Kebudayaan dan Pariwisata Bojonegoro yang mana didapatkan data kunjungan wisata pada tahun 2016-2020.

NO	NAMA DAYA TARIK	2016											
		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEPT	OKT	NOV	DES
1	DANDEL WATERPARK	11760	2455	4079	2964	4760	1843	3776	676	2062	3679	3324	3278
2	KAYANGAN API	39968	16203	2785	4193	1798	11191	8080	2878	2918	3934	1810	4022
3	WADUK PAGAL	4742	1017	1482	1912	2220	4349	3822	1030	1478	1088	2470	2960
4	GROVWOODLAND MGRNUT	958	829	1140	1212	1597	1700	1007	2356	3248	4358	2336	3254
5	KEDUNG MAOR	754	719	836	987	598	478	567	678	551	476	416	780
6	KEDIRI ATAS ANGGAN	41542	2926	3660	4173	1142	26645	2637	1774	3969	1642	1566	10957
7	AGROFISATA BELIDIBESO	2797	3089	3090	10030	13298	11142	18987	8678	7232	3091	8870	15960
8	WISATA EDUKASI GERABAGE	307	1140	2069	1824	1618	1752	337	74	884	1539	988	3579
9	WALI KEDANGAN	123	126	187	253	522	873	182	320	239	432	334	578
10	WISATA MEGAS WONOREJO	31	34	19	35	236	8	0	57	197	254	411	167
11	AGRODUNGA	772	1802	838	1128	998	1807	140	68	250	440	278	684
12	GO FUN	21688	8969	14028	13489	18888	17387	17947	5488	11206	9582	34168	22789
13	AGROFISATA SALAK WEDI	107	28	27	10	11	50	18	803	10	48	30	90
14	WISATA EDUKASI MOODIBESO	39	33	68	40	73	81	59	54	60	79	30	18754
15	PENANGKARAN BUA MALI	31	181	144	433	237	367	96	489	30	117	90	218
16	WISATA SAMEN	26	16	27	27	27	17	27	30	30	25	25	40

Gambar 1 Data Hasil Pre-Processing Kunjungan Wisata Tahun 2016

c. Data Cleaning

Dalam proses ini menghilangkan data atau huruf yang tidak tersedia dan duplikasi data yang menyebabkan data menjadi tidak konsisten dengan cara data cleaning (menghapus data).

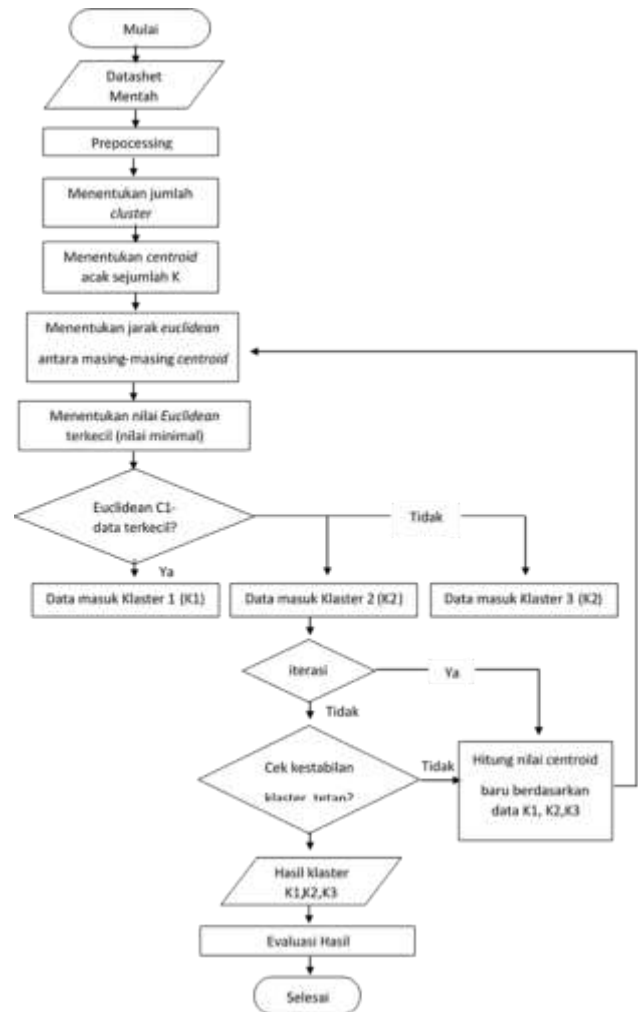


Gambar 2 Diagram Alir Proses Cleaning

d. Proses Clustering Algoritma K-Means

Penghitungan bobot yang menggunakan algoritma K-Means, pertama yang dilakukan yaitu menentukan jumlah kluster dengan pengusulan jumlah 3 cluster yaitu tinggi, sedang dan rendah. Selanjutnya menentukan centroid awal secara acak, setelah memperoleh nilai centroid, kemudian melakukan

penghitungan Euclidean (iterasi) dengan menghitung jarak dataset dengan pusat kluster.



Gambar 3 Metode yang diusulkan

HASIL DAN PEMBAHASAN

Proses implementasi dan pembahasan sesuai flowchat metodologi penelitian ini adalah sebagai berikut:

a. Mempersiapkan data Jumlah Kunjungan Wisata

Dataset jumlah kunjungan wisata diambil dari Kantor Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro dimana data tersebut merupakan data Kunjungan Wisata di Kabupaten Bojonegoro tahun 2016 – 2020. Dalam melakukan clustering,

data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah kunjungan wisata. Data akan diolah dengan melakukan *clustering* jumlah kunjungan wisata dalam 3 *cluster* yakni *cluster* tingkat kunjungan tinggi, sedang, dan rendah. Data jumlah kunjungan wisata dapat dilihat pada gambar 1.

b. Preprocessing

Metode *preprocessing* yang digunakan adalah *missing value*, *inconsistent data*, dan *noisy data* dengan data *cleaning* yaitu data yang masuk pada 3 kategori tersebut akan dihapus. Dari 18 destinasi wisata ada 2 data yang di *cleaning* sehingga tersisa sebanyak 16 destinasi wisata. Sebanyak 16 destinasi wisata tersebut siap untuk diproses kedalam metode atau algoritma. Sample data *preprocessing* dapat dilihat pada gambar berikut.



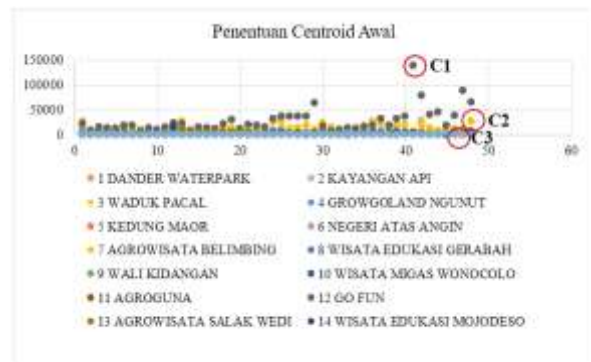
Gambar 4 Sample Data Preprocessing (Gofun)

c. Menentukan Jumlah Cluster

Salah satu cluster yang ingin dikelompokan/dicari yaitu cluster tingkat kunjungan wisata tinggi, cluster tingkat kunjungan sedang dan cluster tingkat kunjungan rendah.

d. Menentukan Centroid Awal

Dalam penerapan algoritma *K-Means* dihasilkan nilai titik tengah atau *centroid* dari data. *Centroid* awal ditentukan secara acak atau bisa menggunakan diagram *scatter* seperti pada diagram gambar 5 dibawah, proses pencarian nilai titik tengah dilakukan dengan mengambil nilai terbesar (maksimum) untuk *centroid* 1 (C1) adalah destinasi wisata Go Fun, nilai rata-rata untuk *cluster centroid* 2 (C2) adalah destinasi wisata agrowisata belimbing dan nilai terkecil (minimum) untuk *cluster centroid* 3 (C3) adalah destinasi wisata samin. Dimana satuan dalam bilangan tersebut adalah kunjungan.



Gambar 5 Penentuan Centroid Awal

Tabel 1 Data Centroid Awal

NAMA WISATA	TAHUN 2016													
	AGB	centroid 1	21668	8569	14025	13469	10886	17267	17647	5486	13106	9202	14116	22789
GF	centroid 2	25292	8000	9000	10150	12456	11345	10987	9876	7234	3690	9870	18960	
WS	centroid 3	26	30	25	27	27	15	25	20	30	25	25	40	
NAMA WISATA	TAHUN 2017													
	AGB	centroid 1	21668	8569	14025	13469	10886	21691	29778	10894	19347	18074	14568	30734
GF	centroid 2	25292	8000	9640	10150	9150	13300	13300	5100	8900	9250	7025	26605	
WS	centroid 3	26	30	25	27	27	15	25	20	39	25	0	31	
NAMA WISATA	TAHUN 2018													
	AGB	centroid 1	36557	36797	36252	36267	61840	15678	5440	8920	12890	10960	15674	17980
GF	centroid 2	26345	10265	12955	17660	9965	22420	14700	8100	10900	12875	10630	25925	
WS	centroid 3	124	135	117	623	165	96	85	72	1155	91	74	66	
NAMA WISATA	TAHUN 2019													
	AGB	centroid 1	30467	17503	31650	36300	137587	78023	39581	43531	16902	88287	88063	63860
GF	centroid 2	22926	10925	17600	19300	5300	20900	12700	7900	9500	11160	9500	25500	
WS	centroid 3	61	103	111	164	4	10	19	11	2065	10	25	520	
NAMA WISATA	TAHUN 2020													
	AGB	centroid 1	0	0	0	0	0	0	0	0	1381	2004	330	0
GF	centroid 2	15350	8200	0	0	0	0	0	0	8800	11000	9900	7500	4800
WS	centroid 3	0	0	0	0	0	0	0	0	0	0	0	0	0



e. Menghitung Euclidean (Iterasi)

1. Iterasi 0

Menghitung jarak *dataset* dengan pusat kluster dengan menggunakan perhitungan jarak *Euclidean Distance* (persamaan 1). Berikut adalah perhitungan jarak *Euclidean Distance* antara *centroid* dan *dataset* hasil preprocessing (lampiran 1-5) data ke-1. Tahun 2016 (Data ke 1 dengan centroid 1)

$$D_{L1}(\chi_2, \chi_1) = |\chi_2 - \chi_1| = \sqrt{\sum_{j=1}^p (\chi_{2j} - \chi_{1j})^2}$$

$$D_{L1}(\chi_2, \chi_1)$$

$$= \sqrt{(11766 - 25292)^2 + (2455 - 8000)^2 + (4079 - 9000)^2 + (5964 - 10150)^2 + (4766 - 12456)^2 + (5845 - 11345)^2 + (5776 - 10987)^2 + (676 - 9876)^2 + (2082 - 7234)^2 + (3679 - 3690)^2 + (3924 - 9870)^2 + (5279 - 18960)^2}$$

$$= \sqrt{(-13526)^2 + (-5545)^2 + (-4921)^2 + (-4186)^2 + (-7690)^2 + (-5500)^2 + (-5211)^2 + (-9200)^2 + (-5152)^2 + (-11)^2 + (-5946)^2 + (-13681)^2}$$

$$= \sqrt{182952676 + 30747025 + 24216241 + 17522596 + 59136100 + 30250000 + 27154521 + 84640000 + 26543104 + 121 + 35354916 + 187169761}$$

$$= \sqrt{705687061}$$

$$= 26,564.77$$

Setelah perhitungan jarak selesai, jarak tersebut dipakai untuk penentuan kluster dari data 1 yaitu didasarkan pada jarak terpendek/terkecil. Berdasarkan jarak antara C1-Data1 : tahun 2016 (34,426.16), C2-Data2 tahun 2016 (26,564.77) dan C3-Data1 : tahun 2016 (18,589.09), maka data 1 tahun 2016 termasuk kluster 3 karena merupakan jarak terpendek.

2. Iterasi 1

Dalam *K-Means*, iterasi pertama pasti akan dilakukan karena digunakan untuk mengecek kestabilan kluster. Pada iterasi pertama, penentuan pusat *centroid*

menggunakan hasil kluster (jarak terpendek) dari iterasi 0 dimana *centroid* iterasi kesatu dihasilkan dari total atribut dibagi banyaknya anggota data pada kluster tersebut.

$$\text{Atribut dander waterpark pada c1 tahun 2016} = \sum \frac{ai}{n} = \frac{21668}{1} = 21668$$

$$\text{Atribut dander waterpark pada c2 tahun 2016} = \sum \frac{ai}{n} = \frac{25292}{1} = 25292$$

$$\text{Atribut dander waterpark pada c3 tahun 2016} = \sum \frac{ai}{n} = \frac{42638}{14} = 3045.571$$

Lakukan seterusnya untuk seluruh atribut dan setiap bulan dan tahunnya dengan masing-masing centroid sehingga menghasilkan centroid seperti table 2.

Tabel 2 Penentuan Centroid Iterasi 1

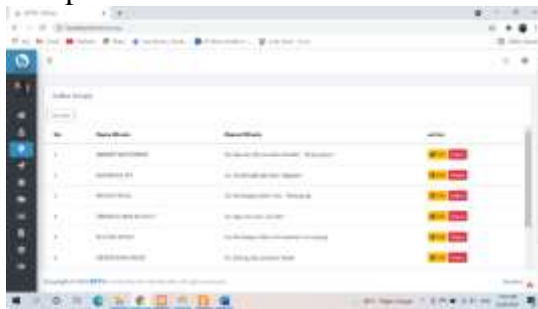
Centroid Tahun 2016												
NAMA	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Ag	Sep	Okt	Nov	Des
WISATA												
AGB centroid 1	21668	8569	14025	13469	10886	17267	17647	5486	13106	9202	14116	22789
GF centroid 2	25292	8000	9000	10150	12456	11345	10987	9876	7234	3690	9870	18960
WS centroid 3	3045.571	1072	1196.143	1502.429	1310.286	2891.714	1982.143	825.3571	998.6429	1339.5	1217.571	3458.071
Centroid Tahun 2017												
WISATA												
AGB centroid 1	21668	8569	14025	13469	10886	21691	29778	10894	19347	18074	14568	30734
GF centroid 2	25292	8000	9640	10150	9150	13300	13300	5100	8900	9250	7025	26605
WS centroid 3	3889.714	1605.357	1560	1851.857	1805.5	1336.386	2104.286	1279.214	1592.429	1257.5	995.8571	2912.571
Centroid Tahun 2018												
WISATA												
AGB centroid 1	36557	36797	36282	36267	61840	15678	5440	8920	12890	10960	15674	17980
GF centroid 2	26345	10265	12955	17660	9965	22420	14700	8100	10900	12875	10630	25925
WS centroid 3	2708.786	1109.214	1319	1387.929	1146.286	2676.429	1870.929	1280.5	1433.286	1516.071	1313.5	2724.357
Centroid Tahun 2019												
WISATA												
AGB centroid 1	30467	17503	31650	36300	137587	78023	39581	43531	16902	38287	88063	63860
GF centroid 2	17311.5	7164	10552	12458	5446	28909.5	8463	5422	6995.5	7881.5	7568.5	18837.5
WS centroid 3	1670.385	968.0769	1083.462	1187.769	176.1538	2465.546	1089.308	487.6154	811.6923	818.9231	678.8462	1638.769
Centroid Tahun 2020												
WISATA												
AGB centroid 1	3148.75	1389.75	0	0	0	0	0	1611.75	1943.25	2949	1600.5	1096.75
GF centroid 2	13875	6167.5	0	0	0	0	0	6148	7680	7219.5	6469.5	3812.5
WS centroid 3	293.6	274.6	0	0	0	0	3.7	32.1	34.4	98.4	74.5	68.4

Lakukan perhitungan jarak *Euclidian Distance* pada *dataset* dengan *centroid* iterasi 1 seperti langkah diiterasi 0.

Pada iterasi pertama dilakukan cek kestabilan terkait hasil kluster iterasi 0 dan iterasi 1. Iterasi akan dilanjutkan lagi jika ada data yang berpindah kluster dari hasil iterasi 0 ke iterasi 1 atau dihentikan jika hasil kluster sama atau stabil. Pada iterasi 1, data yang tetap klasternya sebanyak 16 data pada tahun 2016-2018 dan data yang berpindah kluster sejumlah 1 data pada tahun 2019-2020. Tabel 3 memperlihatkan hasil pengecekan kestabilan kluster per tahunnya antara iterasi 0 dan iterasi 1.



4. Tampilan Halaman Destinasi Wisata

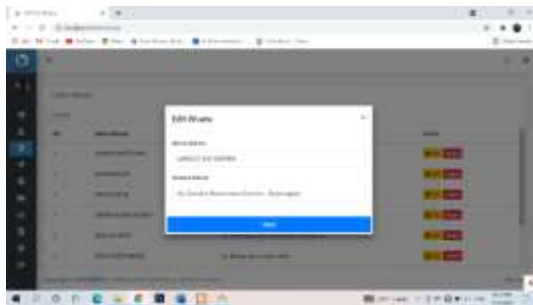


Gambar 9 Halaman Destinasi Wisata



Gambar 13 Halaman Daftar Kunjungan Wisata

6. Tampilan Halaman Penentuan Centroid



Gambar 10 Halaman Destinasi Wisata



Gambar 10 Halaman Destinasi Wisata



Gambar 14 Halaman Penentuan Centroid

5. Tampilan Halaman Daftar Kunjungan Wisata



Gambar 12 Halaman Daftar Kunjungan
Wisata

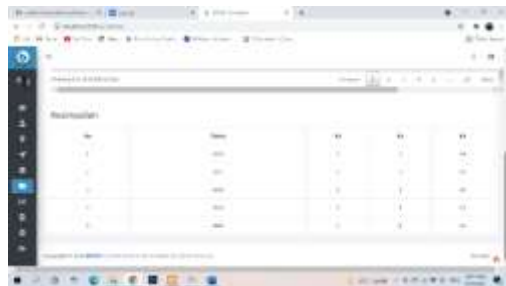


Gambar 15 Halaman Penentuan Centroid

7. Tampilan Halaman Klaster Destinasi Wisata

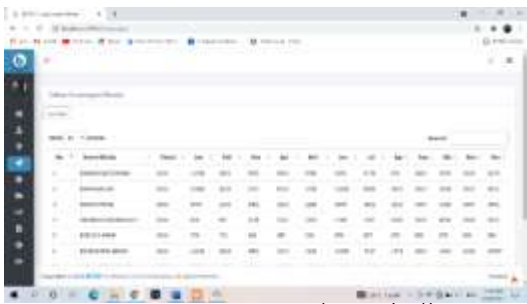


Gambar 16 Halaman Klasterisasi

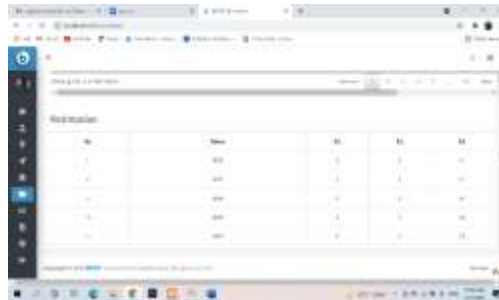


Gambar 16 Halaman Klasterisasi

8. Tampilan Halaman Grafik Kunjungan Wisata

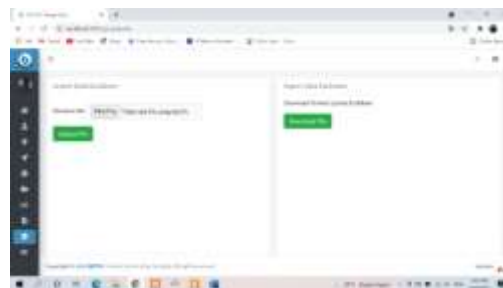


9. Tampilan Halaman Laporan Wisata



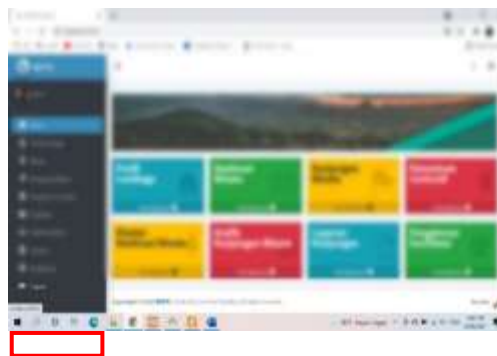
Gambar 19 Halaman Laporan

10. Tampilan Halaman Pengaturan



Gambar 20 Halaman Pengaturan Data

11. Tampilan Halaman Logout



Gambar 21 Halaman Menu Logout

PENUTUP

Kesimpulan

Kesimpulan yang dapat diambil dari hasil Analisa dan komputasi penerapan *unsupervise learning* untuk klasterisasi jumlah kunjungan wisata di Bojonegoro dengan metode *K-Means* perhitungan jarak *Euclidean Distance* adalah sebagai berikut :



- a) Dapat diketahui cara perhitungan klusterisasi kunjungan wisata menggunakan Implementasi *Unsupervise Learning* dengan metode *Clustering* dan Algoritma *K-Means* yang telah dilakukan pada 3 usulan klaster yang terdiri dari klaster tinggi, sedang dan rendah. Didapatkan hasil anggota klaster kunjungan wisata dengan anggota C1 pada tahun 2016-2019 sejumlah 1 data, tahun 2020 sejumlah 3 data. Anggota C2 pada tahun 2016-2018 sejumlah 1 data, tahun 2019 sejumlah 3 data dan tahun 2020 sejumlah 2 data. Sedangkan jumlah anggota klaster kunjungan rendah pada tahun 2016-2018 sejumlah 14 data, tahun 2019 sejumlah 12 data dan tahun 2020 sejumlah 11 data.
- b) Berdasarkan penelitian dapat diketahui tingkat kecondongan kepadatan pengunjung wisata menggunakan implementasi *Unsupervise Learning* metode *Clustering* dan Algoritma *K-Means* dengan hasil iterasi stabil pada iterasi 3. Dapat disimpulkan bahwa tingkat kepadatan pengunjung wisata dengan klaster tinggi pada tahun 2016-2019 lebih condong kepada wisata GoFun (data ke 12), sedangkan pada tahun 2020 tingkat kepadatan pengunjung wisata dengan klaster tinggi lebih condong pada wisata Kayangan Api (data ke 2), Waduk Pacal (data ke 3) dan Growgoland Ngunut (data ke 4). Tingkat kepadatan pengunjung wisata dengan klaster sedang tahun 2016-2018 lebih condong pada wisata Agrowisata Belimbing (data ke 7), tahun 2019 tingkat kepadatan pengunjung klaster sedang yaitu Dander Waterpark (data ke 1), Kayangan api (data ke 2) dan Agrowisata Belimbing (data ke 7). Sedangkan Klaster kunjungan wisata rendah pada tahun 2016-2018 lebih condong pada data ke 1,2,3,4,5,6,8,9,10,11,13,14,15,16 dan tahun 2019 terdapat pada data ke 3,4,5,6,8,9,10,11,13,14,15,16 sedangkan tahun 2020 terdapat pada data ke 5,6,8,9,10,11,12,13,14,15,16.
- c) Sistem BOTIV dibangun dengan menggunakan *framework Codeigniter*, keuntungan yang didapat yaitu memudahkan dalam pengembangan lebih lanjut oleh *developer* yang ingin berkontribusi dikemudian hari. Sistem Klasterisasi dengan menggunakan algoritma *K-Means* untuk mengetahui tingkat kepadatan pengunjung wisata di Bojonegoro sudah didukung oleh *security access* dimana pihak yang tidak diperbolehkan dilarang mengakses bagian administrator tanpa izin pihak berwenang. Sistem Klasterisasi sudah didukung dengan tampilan *responsive* menggunakan material *design* artinya apabila *website* sudah menggunakan domain bisa diakses oleh admin dan user yang berkepentingan. Sistem Klasterisasi dengan menggunakan algoritma *K-Means* untuk mengetahui tingkat kepadatan pengunjung wisata di Bojonegoro menggunakan basis data MYSQL sebagai tempat penyimpanan data yang terkait dalam jumlah kunjungan wisata Bojonegoro yang berada dilingkungan Dinas Kebudayaan dan Pariwisata Bojonegoro. Setelah diakumulasi, nilai yang didapat dari angket uji kelayakan dengan skala penilaian 1 sampai 4 adalah sebanyak 1308 dari total 1440 atau 90,8% dari *test engineering* menyatakan bahwa Sistem Klasterisasi BOTIV (Bojonegoro Toursm Visit) “Valid” (baik) dan bisa digunakan.

Saran

Pada hasil penelitian ini, saran yang diajukan berdasarkan kesimpulan selama pembangunan Sistem Klasterisasi dengan menggunakan algoritma *K-Means* untuk mengetahui tingkat kepadatan pengunjung wisata di Bojonegoro ini sebagai berikut:

- a) Dalam penelitian ini, perhitungan jarak anatara data dan *centroid* hanya diuji menggunakan *Euclidean Distance*.



- Diperlukan uji metode jarak lain seperti *Manhattan* dan *Minkowski*.
- b) Perlunya dilakukan perawatan secara berkala pada web Sistem “BOTIV” Klasterisasi dengan menggunakan algoritma *K-Means* untuk mengetahui tingkat kepadatan pengunjung wisata di Bojonegoro yang sudah dibangun. Hal ini dimaksudkan untuk menghindari terjadinya kesalahan-kesalahan yang muncul diakibatkan oleh fungsi-fungsi yang mungkin saja mengalami error.
- [8] B. Nurhayati and R. P. Iswara, “Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial sebagai media promosi Online Bagi Masyarakat,” *J. Tek. Inf.*, vol. 12, no. 1, pp. 79–96, 2019.

DAFTAR PUSTAKA

- [1] Redaksi, “Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro.” <https://idalamat.com/alamat/24914/dinas-kebudayaan-dan-pariwisata-kabupaten-bojonegoro>.
- [2] I. Pramudiono, “Apa itu data mining,” *Penerbit Andi, Yogyakarta*, 2006.
- [3] R. W. Sari and D. Hartama, “Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi,” in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2018, vol. 1, no. 1.
- [4] E. Irwansyah and M. Faisal, *Advanced Clustering: Teori dan Aplikasi*. Deepublish, 2015.
- [5] L. Teori, “(K-MEANS ALGORITHM IMPLEMENTATION FOR CLUSTERING OF PATIENTS DISEASE IN KAJEN CLINIC OF PEKALONGAN) Anindya Khrisna Wardhani Magister Sistem Informasi Universitas Diponegoro,” vol. 14, pp. 30–37, 2016.
- [6] Flinsetyadiz, “Memahami Konsep Data Mining Beserta Prosesnya,” 2020. <https://flinsetyadi.com/memahami-konsep-data-mining-beserta-prosesnya/>.
- [7] C. Astria, A. P. Windarto, A. Wanto, and E. Irawan, “Metode K-Means pada Pengelompokan Wilayah Pendistribusian Listrik,” *Semin. Nas. Sains Teknol. Inf.*, pp. 306–312, 2019.