Implementasi Data Mining Dalam Klasifikasi Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Algoritma K- Means

# Bambang Setio Purnomo1, Putri Taqwa Prasetyaningrum2

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta

Yogyakarta, Indonesia

*Email : 1**bambang.15121016@gmail.com,* *2putri@mercubuana-yogya.ac.id*

Diajukan : 14 februari 2020

***Abstrak***

*Dari banyaknya data kunjungan wisatawan pada setiap bulan dan tahunnya sehingga sangat sulit untuk mengetahui banyaknya jumlah wisatawan yang terjadi pada setiap bulannya. Pada saat peneliti melakukan kerja praktek di Dinas Pariwisata problem seperti ini yang masih menjadi PR bagi Dinas Pariwisata untuk mengolah data jumlah kunjungan wisatawan. Dari pemikiran di atas maka, dilakukan peneliti tentang klasifikasi Kunjungan Wisatawan Di Kota Yogyakarta menggunakan metode K-Means. Tujuannya adalah untuk mengetahui objek wisata yang memiliki potensi yang paling rendah dalam kunjungan wisatawan sampai yang paling tinggi. Klasifikasi tersebut dapat menggunakan metode Algoritma K-Means karena metode ini 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). Dari data yang diinputkandan telah di proses melalui metode Algoritma K-Means bahwa telah melakukan iterasi sebanyak 5 kali dengan memilih cluster 1, cluster 2, cluster 3 secara acak (random) dengan cluster 1 memiliki 24 data dengan persentase sebesar (50%), cluster 2 memiliki 11 data dengan persentase*

*sebesar (23%), dan cluster 3 memiliki 13 data dengan persentase sebesar (27%).*

***Kata kunci:*** *klasifikasi kunjungan wisatawan, Data Mining, Algoritma K-Means, Knowledge Discovery In Databases (KDD), dinas pariwisata.*

***Abstract***

*From a large number of tourist visits data every month and year, it is difficult to know the number of tourist visits each month. When the researcher did an internship in the Department of Tourism, this kind of problem still become unfinished work for the Department of Tourism to process data on the number of tourist visits. From the above thought, a study was conducted on the classification of Tourist Visits in Yogyakarta City using the K-Means method. The aim is to find out which tourist attractions have the lowest potential of visits to the highest. The classification can use the K-Means Algorithm method because this method is very easy to implement and uses simple principles and can be explained in non-statistics, and it is very suitable for unsupervised research. From the data entered and processed through the K-Means Algorithm method, that iterated 5 times by selecting cluster 1, cluster 2, cluster 3 randomly with cluster 1 having 24 data with a percentage of (50%), cluster 2 has 11 data with a percentage of (23%), and cluster 3 has 13 data with a percentage of (27%).*

***Keywords:*** *tourist visit classification, Data Mining, K-Means Algorithm, Knowledge Discovery In Databases (KDD), tourism department.*

# Pendahuluan

Yogyakarta merupakan salah satu kota di Indonesia yang memiliki daya tarik wisata dan merupakan kota tujuan wisata yang paling diminati oleh wisatawan, dilihat dari jumlah kunjungan wisatawan yang semakin naik dari tahun ke tahun. Selain sebagai kota wisata, Yogyakarta merupakan kota pelajar, kota budaya dan kota perjuangan. Karena Yogyakarta disebut sebagai kota wisata, banyak berbagai macam objek wisata yang ditawarkan oleh Kota Yogyakarta. Mulai dari wisata budaya, wisata alam, wisata religi, wisata kuliner, wisata minat khusus dan wisata belanja. Wisata budaya merupakan salah satu objek wisata yang paling diminati oleh wisatawan jika berkunjung ke Yogyakarta. Pertumbuhan tingkat kunjungan wisatawan di kota yogyakarta tahun 2014-2018 sejumlah 2014 sebanyak 3.007.253 wisatawan, 2015 sebanyak 3.250.681 wisatawan, 2016 sebanyak

3.261.748 wisatawan, 2017 sebanyak 3.894.711 wisatawan, dan 2018 sebanyak 4.103.240 wisatawan (Sumber : Dinas Pariwisata Kota Yogyakarta, 2018).

Dari banyaknya data kunjungan wisatawan pada setiap bulan dan tahunnya sehingga sangat sulit untuk mengetahui banyaknya jumlah wisatawan yang terjadi pada setiap bulannya. Pada saat peneliti melakukan *kerja praktek* di Dinas Pariwisata problem seperti ini yang masih menjadi *PR* bagi Dinas Pariwisata untuk mengolah data jumlah kunjungan wisatawan.

Dari pemikiran di atas maka, dilakukan peneliti tentang klasifikasi Kunjungan Wisatawan Di Kota Yogyakarta menggunakan metode *K-Means*. Tujuannya adalah untuk mengetahui objek wisata yang memiliki potensi yang paling rendah dalam kunjungan wisatawan sampai yang paling tinggi. Mengkelompokan tersebut dapat menggunakan metode Algoritma *K-Means* karena metode ini 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 datamining mampu menjadi solusi dalam menganalisa data. Dengan data yang sudah dikelompokkan menggunakan Algoritma *K-Means* diharapkan dapat mempermudah Dinas Pariwisata Kota Yogyakarta dalam klasifikasi hasil jumlah kunjungan wisatawan.

# Metode Penelitian

Penelitian ini dilakukan dengan menggunakan beberapa tahan sesuai yang dijelaskan pada gambar 1, seperti pengumpulan data, *pre-processing* data, pembangunan sistem menggunakan metode *K-Means*.

*Knowledge Discovery*

*In Databases (KDD)*

Gambar 1. *Flowcat* proses mengolah data dengan Algoritma *K-Means*

# Pengumpulan Data

Pengumpulan data dilakukan untuk memenuhi kebutuhan informasi data yang digunakan dalam penelitian. Pada tahap ini penelitian mendapatkan data dari Dinas Pariwisata Kota Yogyakarta dari tahun 2014-2018 dengan 48 data yang memiliki 3 atribut seperti : nama objek wisata, nama bulan, nama tahun.

# Pre-Processing Data

data yang di dapat dari Dinas Pariwisata Kota Yogyakarta merupakan data mentah yang belum siap di olah, leh karena itu perlu dilakukan tahap *pre-processing* untuk mendapatkan data yang siap untuk diolah pada proses selanjutnya. pertambangan informasi di database tidak teratur seperti pertambangan informasi adalah bagian dari prosedur pengungkapan *knowledge* dalam database itu sendiri *(KDD),* meskipun begitu pula informasi pertambangan adalah prosedur yang paling penting selama ini pengungkapan menghabiskan informasi dalam database *(KDD)* (han and kamber, 2006). Tahap *Pre-Processing* yang dilakukan adalah

* + 1. Seleksi informasi, *determinasi* informasi dari suatu susunan informasi operasional harus dilakukan sebelum *fase extricating* data dalam *KDD* dimulai. Informasi dari penentuan yang akan digunakan untuk proses penambangan informasi, menyingkirkan dalam dokumen terpisah dari database operasional.
		2. *Pre-penganan* / pembersihan, sbelum prosedur pertambangan informasi dapat dilakukan, proses pembersihan penting pada informasi yang menjadi pusat *KDD*. Proses pembersihan dapat menggabungkan salinan informasi yang saling bertentangan, dan menyesuaikan kesalahan dalam informasi, misalnya, kesalahan pencetakan.
		3. Perubahan, *coding* adalah prosedur perubahan pada informasi yang telah dipilih, sehingga informasi yang membandingkan peroses penambangan informasi. *KDD* proses *coding* di proses *inventif* dan sangat tunduk pada jenis atau contoh data yang akan tampak dalam *database*.
		4. Informasi pertambangan, pertambangan informasi adalah jalan menuju mencari contoh atau data menarik dalam memilih infomasi dengan memanfaatkan prosedur atau teknik strategi, teknik, atau perhitungan pertimbangan informasi dalam sangat bervariasi. Pilihan strategi yang tepat atau perhitungan tergantung *intens* pada tujuan dan proses *KDD* umumnya.
		5. Memahami / evaluasi, data *desain* yang dihasilkan dari prosedur informasi pertambangan harus ditampilkan dalam struktur efektif dipahami oleh mitra. Tahap ini adalah bagian dari prosedur *KDD* disebut pemahan. Tahap ini mengabungkan memeriksa apakah contoh atau data ditemukan ditolak oleh kebenarannya atau teori-teori yang ada beberapa waktu baru- baru ini (han, 2006).
		6. *Knowledge*, Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

# Clustering Data Menggunakan Algoritma K-Means

Metode *clustering* menurut kategori kekompakan terbagi menjadi dua yaitu *komplet dan parsial*. Semua data bisa dikatakan kompak menjadi satu kelompok jika semua data bisa bergabung menjadi satu *(dalam konteks penyekatan)* namun jika ada sedikit data yang tidak ikut bergabung dalam kelompok mayoritas data tersebut dikatakan mempunyai perilaku menyimpang. Data yang menyimpang ini dikenal dengan sebutan *noise*. Metode yang tangguh untuk mendeteksi *noise* ini adalah *DBSCAN* (Wahyu, 2017).

Clustering termasuk ke dalam *descriptive methods*, dan juga termasuk *unsupervised learning* dimana tidak ada pendefinisian kelas objek sebelumnya. Sehingga clustering dapat digunakan untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya (Pada et al., 2018).

Beberapa teknik *clustering* yang paling sederhana dan umum adalah *clustering K-Means*. Secara detail teknik ini menggunakan ukuran ketidak miripan untuk klasifikasi obyek. Ketidak miripan dapat diterjemahkan dalam konsep jarak. Dua obyek dikatakan mirip jika jarak dua objek tersebut dekat. Semakin tinggi nilai jarak, semakin tinggi nilai ketidak miripannya (Handoko, 2016).

Dalam penelitian ini metode yang digunakan yaitu metode data mining sebagai berikut. (a) Tahap pengumpulan data, (b) Tahap pengolahan data, (c) Tahap *Clustering* dan (d) Tahap Analisis. Dapat dilihat digambar 2 berikut adalah *flowchart* Algoritma *K-Means* (Windarto,2017).



Gambar 2. Algoritma *K-Means*

Langkah-langkah yang dipergunakan dalam metode *K-Means* sesuai gambar *flowchart* yang ditampilkan pada gambar 1 adalah sebagai berikut (Sani, 2018):

1. Tentukan jumlah *cluster* yang ingin dibentuk dan tetapkan pusat *cluster* k.
2. Menggunakan jarak euclidean kemudian hitung setiap data ke pusat *cluster.*
3. Kelompokkan data ke dalam *cluster* dengan jarak yang paling pendek dengan persamaan.



1. Hitung pusat *cluster* yang baru menggunakan persamaan.

Dengan : Xij\_ Kluster ke k p = banyaknya anggota *cluster* ke – k

1. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain (Sadewo et al, 2017).

# Hasil Dan Pembahasan

1. *Iterasi* ke 1

a Penentuan pusat *(centroid)* awal *cluster*

Untuk menentukan pusat *(centroid)* awal ditentukan dengan mengacak *(random)* dari data nilai yang sudah ada. Pada kasus ini pusat *(centroid)* awal dapat dilihat di tabel 1.

Tabel 1. Pusat *(Centroid)* Awal *Cluster*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Centroid 1 | 34125 | 29254 | 31536 | 36741 | 38734 |
| Centroid 2 | 53492 | 27354 | 29245 | 40545 | 45626 |
| Centroid 3 | 11507 | 9129 | 9024 | 10247 | 11034 |

1. Perhitungan jarak dengan pusat *cluster* dapat dilihat ditabel 2 dengan rumus : d (x,y) = ‖𝑥−𝑦‖ = √Σ(𝑥𝑖−𝑦𝑖)𝑛𝑖=12 ; i= 1,2,3,…n

Perhitungan jarak dari data ke-1 terhadap pusat *cluster* adalah : C(1,1)=√( 53492−34125)2+( 27354−29254)2+(29245−31536)2+

(40545−36741)2+(45626−38734)2 = 2111657

C(1,2)=√( 53492−53492)2+( 27354−27354)2+(29245−29245)2+

(40545−40545)2+(45626−45626)2 = 0

C(1,3)=√( 53492−11507)2+( 27354−9129)2+(29245−9024)2+

(40545−10247)2+(45626−11034)2 = 6795848

Tabel 2. Perhitungan jarak dengan pusat *cluster*

|  |
| --- |
| **iterasi 1** |
| **data ke** | **jarak c1** | **jarak c2** | **jarak c3** | **cluster terdekat** |
| 1 | 21116,57 | 0 | 67958,48 | 2 |
| 2 | 8520,885 | 25967,84 | 49155,92 | 1 |
| 3 | 21298,04 | 15087,45 | 57633,51 | 2 |
| 4 | 8196,684 | 18193,44 | 56780,07 | 1 |
| 5 | 68193,36 | 49540,56 | 104529,8 | 2 |
| 6 | 44892,96 | 35687,02 | 68813,56 | 2 |
| 7 | 25056,91 | 43861,74 | 50470,07 | 1 |
| 8 | 8264,558 | 26701,26 | 46896,35 | 1 |
| 9 | 14842,45 | 33311,51 | 40303,3 | 1 |
| 10 | 0 | 21116,57 | 53782,91 | 1 |

Keterangan :

* + C1 memiliki 21 anggota, yaitu data ke-2, data ke-4, data ke-7, data ke-8, data ke-9, data ke-10, data ke-11, data ke-12, data ke-13, data ke-14, data ke-15, data ke-16, data ke-17, data ke-18, data ke-19, data ke-20, data ke-21, dan data ke-22 data ke-23, data ke-25, dan data ke-26.
	+ C2 memiliki 15 anggota, yaitu data ke-1, data ke-3, data ke-5, data ke-6, data ke-24, data ke-27, data ke-28, data ke-29, data ke- 30, data ke-31, data ke-32, data ke-33, data ke-34, data ke-35, dan data ke-36.
	+ C3 memiliki 12 anggota, yaitu data ke-37, data ke-38, data ke-39, data ke-40, data ke-41, data ke-42, data ke-43, data ke-44, data ke- 45, data ke-46, data ke-47, dan data ke-48.
1. Perhitungan pusat *cluster* baru.

Setelah di ketahui anggota tiap-tiap *cluster* kemudian pusat *cluster* baru di hitung berdasarkan data anggota tiap-tiap *cluster* sesuai dengan rumus pusat anggota *cluster* yaitu :𝑣=Σ𝑥𝑖𝑛𝑖=1𝑛 ; i=1,2,3,…n.

Sehingga setiap data yang masuk ke dalam setiap *cluster* maka di jumlah lalu di bagi dengan banyaknya data yang masuk kedalam *cluster* tersebut dapat dilihat di gambar 2.



Gambar 2. Perhitungan pusat *cluster* baru

C1

* + 507804:21 = 241811
	+ 851950:21 = 40569
	+ 928009:21 = 441909
	+ 984173:21 = 468654
	+ 1071535:21 = 510255

C2

* + 1062712:15 = 708475
	+ 856221:15 = 570814
	+ 829330:15 = 552887
	+ 892425:15 = 59495
	+ 1080266:15 = 720177

C3

* + 92439:12 = 770325
	+ 142560:12 = 11880
	+ 139439:12 = 116199
	+ 159421:12 = 132851
	+ 181154:12 = 150962

Hasil pembentukan pusat *(centroid)* baru dapat dilihat di tabel 3.

Tabel 3. Hasil Pembentukan Pusat *Cluster* Baru

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Centroid 1 | 241811 | 40569 | 441909 | 468654 | 510255 |
| Centroid 2 | 708475 | 570814 | 552887 | 59495 | 720177 |
| Centroid 3 | 770325 | 11880 | 116199 | 132851 | 150962 |

Setelah di lakukan iterasi selama 5 (lima) kali diperoleh nilai seperti pada tabel 3 dan dikarenakan anggota *centroid* 1, *centroid* 2, *centroid* 3 pada *iterasi* ke 5 (lima) tidak mengalami perubahan, maka *iterasi* dihentikan.

Tabel 3. Hasil dari perhitungan Algoritma *K-Means*

|  |
| --- |
| **iterasi 5** |
| **data ke** | **Jarak c1** | **Jarak c2** | **Jarak c3** | **Cluster** |
| 1 | 25115,22 | 79886,6 | 63709,79 | cluster 1 |
| 2 | 22150,94 | 93200,31 | 42925,35 | cluster 1 |
| 3 | 32111,32 | 92526,88 | 54404,65 | cluster 1 |
| 4 | 18327,12 | 86515,3 | 51532,63 | cluster 1 |
| 5 | 68711,36 | 87345,89 | 102912,8 | cluster 1 |
| 6 | 53141,84 | 102923 | 67862,6 | cluster 1 |
| 7 | 30942,24 | 98535,97 | 42721,56 | cluster 1 |
| 8 | 23627,24 | 95032,14 | 40951,32 | cluster 1 |
| 9 | 29844,21 | 101869,3 | 34263,32 | cluster 1 |
| 10 | 17363,73 | 88283,63 | 47917,34 | cluster 1 |

**Persebaran Data**

**23%**

**50%**

Cluster 1 SEDANG

Cluster 2 TINGGI

Cluster 3 RENDAH

**27%**

Hasil grafik dari persebaran data dengan menggunakan metode Algoritma *K-Means* bahwa telah melakukan *iterasi* sebanyak 5 kali dengan memilih *cluster* 1, *cluster* 2, *cluster* 3 secara acak (random) dengan *cluster* 1 memiliki 24 data dengan *persentase* sebesar (50%), *cluster* 2 memiliki 11 data dengan *persentase* sebesar (23%), dan *cluster* 3 memiliki 13 data dengan *persentase* sebesar (27%).

# Kesimpulan Dan Saran

Data kunjungan wisatawan Kota Yogyakarta bisa kita lihat dari 3 *cluster* yaitu *cluster* 1, *cluster* 2, *cluster* 3. *Cluster* 1 yakni jumlah kunjungan wisatawan dengan kategori sedang dengan rata-rata 15.611 wisatawan sampai 46.783 wisatawan dengan 24 data dan *persentase* sebesar (50%) masuk kedalam objek wisata kraton pada bulan januari sampai bulan desember, taman pintar pada bulan januari, februari, april, mei, juni, juli, september, oktober, november, dan di objek wisata gembira loka zoo pada bulan agustus, september, oktober. *Cluster* 2 yakni jumlah kunjungan wisatawan dengan kategori tinggi dengan rata-rata 46.784 wisatawan sampai 91.566 wisatawan dengan memiliki 11 data dengan *persentase* sebesar (23%) masuk kedalam objek wisata taman pintar pada bulan maret dan desember, dan di objek wisata gembira loka zoo pada bulan januari, februari, maret, april, mei, juni, juli, november, dan desember. *Cluster* 3 jumlah kunjungan wisatawan dengan kategori rendah 0 wisatawan sampai 15.610 wisatawan memiliki 13 data dengan *persentase* sebesar (27%) masuk kedalam objek wisata taman pintar pada bulan agustus, dan di objek wisata museum pada bulan

januari sampai desember. Semoga dengan dilakukannya penelitian ini dapat membantu pihak Dinas Pariwisata Kota Yogyakarta dalam mengolah data kunjungan wisatawan.

# Daftar Pustaka

Adiya, M. H., & Desnelita, Y. (2019). *Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru*. *01*, 17–24.

Handoko, K. (2016). *PENERAPAN DATA MINING DALAM MENINGKATKAN MUTU PEMBELAJARAN PADA INSTANSI PERGURUAN TINGGI MENGGUNAKAN METODE K - MEANS CLUSTERING ( STUDI KASUS DI PROGRAM STUDI TKJ AKADEMI KOMUNITAS SOLOK SELATAN )*. *02*(03), 31–40.

Masruro, A., & Luthfi, E. T. (2014). *SISTEM PENUNJANG KEPUTUSAN PENENTUAN LOKASI WISATA MENGGUNAKAN K-MEANS CLUSTERING DAN TOPSIS Abstraksi Pendahuluan*

*Tinjauan Pustaka*. *15*(04), 1–5.

Pada, T., Hairos, C. V, & Menggunakan, I. (2018). *Penerapan Data Mining Untuk Pengelompokan Wahana*. *17*(1), 92–104.

Wahyu, mikael aditya. (2017). *Plagiat merupakan tindakan tidak terpuji penerapan metode k- means clustering untuk mengelompokan potensi produksi buah – buahan di provinsi daerah istimewa yogyakarta*.

Maulida, L., Studi, P., & Informatika, M. (2018). *KUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DI PROV . DKI JAKARTA DENGAN K-MEANS*. *2*(3), 167–174.

Metisen, B. M., & Sari, H. L. (2015). *ANALISIS CLUSTERING MENGGUNAKAN METODE K- MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA*. *11*(2), 110–118.

Data Jumlah Kunjungan Wisatawan Kota Yogyakarta, Dinas Pariwisata Kota Yogyakarta Tahun 2014 - 2018.

Data Kunjungan Wisatawan Ke Objek Wisata Dari Pengelola Objek Wisata Dan Dinas Kebudayaan Diy Tahun 2014-2018

Sari, R. W., & Hartama, D. (2018). *Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi*. 322–326.

Sari, Y. A., & Dewi, R. K. (2018). *Rekomendasi Lokasi Wisata Kuliner Menggunakan Metode K- Means Rekomendasi Lokasi Wisata Kuliner Menggunakan Metode K-Means Clustering Dan Simple Additive Weighting*. *July*.

Amoes Martin, Hamdani (2018). SISTEM PENDUKUNG KEPUTUSAN UNTUK PENENTUAN TUJUAN WIATA HOTEL DAN ALAM DI INDONESIA.

Yulia, W. (2019).*PENERAPAN DATAMINING UNTUK MENDUKUNG STRATEGI PROMOSI UNIVERSITAS MERCU BUANA YOGYAKATA FAKULTAS TEKNOLOGI INFORMASI.*

Putri ,T,W.(2019) *GRADUATE STUDENT BEST PREDICTION ANALYSIS USING MULTILAYER PERCEPTRON(MLP).*

Syafrudin.(2015).C*ONTOH PERHITUNGAN MANUAL PENERAPAN METODE K-MEANS KLASTERING DATAMINING.*

*Narwati.(2017).Pengelompokkan Mahasiswa Dengan Metode Algoritma K-Means.*