

Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi

Riyani Wulan Sari¹, Dedy Hartama²

¹ Prodi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

² Prodi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: riyaniwulansari24@gmail.com

Abstrak

Perkembangan industri perhotelan bintang di Indonesia cenderung meningkat. Perkembangan dunia perhotelan dalam upaya penyediaan jasa akomodasi pariwisata di Indonesia semakin hari semakin menunjukkan perkembangan yang pesat. Dari berbagai persaingan perhotelan melalui pengelompokan, hotel yang kalah saing dengan hotel yang lain akan memunculkan ide kreatif yang mampu mengangkat derajat hotel yang masih jauh dari persaingan. Kurangnya informasi serta pengelompokan menyangkut kunjungan wisatawan ke berbagai obyek daya tarik wisata di hotel-hotel berbintang di berbagai provinsi mengakibatkan adanya kesulitan bagi Pemerintah Daerah khususnya Dinas Pariwisata dan Ekonomi Kreatif dalam perencanaan strategi pemasaran pariwisata di berbagai provinsi sehingga promosi atau pemasaran pariwisata menjadi tidak tepat sasaran, menyebabkan wisata-wisata di provinsi-provinsi yang kurang dikenal jarang di kunjungi oleh wisatawan asing. Adapun tujuan yang ingin dicapai dalam penelitian ini adalah mendapatkan pengelompokan dari data kunjungan wisatawan di hotel berbintang pada setiap provinsi-provinsi selama kurun waktu tahun 2003-2016 menggunakan Algoritma *K-Means*. *K-Means* merupakan salah satu metode data klustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster / kelompok. Sehingga hasil yang diperoleh dapat menjadi pertimbangan pemerintah khususnya Dinas Pariwisata dan Ekonomi Kreatif di Provinsi-provinsi untuk mengembangkan potensi objek daya tarik wisata di provinsi-provinsi serta sebagai bahan pertimbangan pengembangan hotel-hotel berbintang yang ada di Indonesia.

Kata Kunci: Data Mining, Wisata Asing, Clustering, K-means

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang kaya akan keindahan wisata alam, taman wisata dan taman budaya yang dapat mendukung perkembangan pariwisata di Indonesia. Kondisi letak geografis Indonesia yang kaya akan keindahan alam menjadikan wisatawan asing tertarik untuk mengunjungi Indonesia untuk melihat keindahan alam tersebut [1],[2]. Maka dengan pengelolaan yang baik dan terarah diharapkan pariwisata di Indonesia mampu menarik wisatawan asing untuk berkunjung. Mengingat potensi penduduk Indonesia yang sangat besar karena Indonesia merupakan negara yang memiliki penduduk Muslim terbesar di dunia, yaitu sekitar 207.176.162 juta atau sekitar 88% dari total penduduk pada tahun 2010 oleh BPSRI [3]. Hal itu menjadikan pendapatan negara bertambah dan juga besarnya minat wisatawan berkunjung ke Indonesia memicu meningkatkan peluang dalam jenis berbisnis baik objek wisatanya itu sendiri maupun usaha-usaha lain yang berhubungan dengan aktivitas wisata seperti akomodasi hotel, transportasi, kuliner dan lain sebagainya. Widyawati (2011: 74) mengatakan industri pariwisata terdapat berbagai jenis bisnis yang menunjang pariwisata yaitu salah satunya bisnis perhotelan.

Hotel merupakan salah satu perusahaan jasa yang mengutamakan pelayanan terhadap *customer*. Peraturan Pemerintah Republik Indonesia No. 65 Tahun 2011 tanggal 31 September 2001 Pasal 1, "Hotel adalah bangunan yang khusus disediakan bagi orang untuk dapat menginap atau istirahat, memperoleh pelayanan dan atau fasilitas lainnya dengan dipungut bayaran, termasuk bangunan lainnya yang menyatu dikelola dan dimiliki oleh pihak yang sama kecuali untuk pertokoan dan perkantoran". Kenyamanan pengunjung hotel dalam menggunakan jasa hotel merupakan persyaratan utama, sehingga para pengelola hotel harus menciptakan kenyamanan sebaik mungkin agar jasa hunian hotel dapat dipertahankan. Perkembangan industri perhotelan bintang di Indonesia cenderung meningkat.

2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining juga 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 [4].

2.2 Clustering

Analisis Pengelompokan/*Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain [5].

2.3 K-means

K-means merupakan suatu algoritma yang digunakan dalam pengelompokan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – beda. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya*. [6]. Kemudian algoritma *K-means* akan menguji masing – masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat *cluster*. Selanjutnya posisi pusat *cluster* akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap – tiap *cluster* dan terakhir akan terbentuk *cluster* baru[7]

2.4 Wisatawan Asing

Menurut undang – undang No. 10 tahun 2009 tentang kepariwisataan, disebutkan wisatawan adalah orang yang melakukan wisata. Sedangkan wisata adalah bepergian secara bersama-sama dengan tujuan untuk bersenang-senang, menambah pengetahuan dan lain lain. Sehingga dapat diartikan wisatawan adalah orang-orang yang datang berkunjung pada suatu tempat atau negara yang bukan tempat tinggalnya untuk tujuan tertentu,

3. ANALISA DAN PEMBAHASAN

Hasil pengujian metode Data Mining diuji pada data BPS. Pada penelitian ini perhitungan menggunakan Metode K-Menas[8]. Berdasarkan Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu berdasarkan hasil rata-rata jumlah Tamu Asing pada Hotel Bintang menurut Provinsi[9]. Untuk menyelesaikan penelitian ini dengan menggunakan Metode K-means dalam mengelompokkan data dalam klasifikasi tertentu, menghitung data tersebut[10]

Data tersebut kemudian diakumulasikan dan diambil nilai rata-rata nya, yaitu :

Tabel 1. Nilai rata-rata setelah data alternatif diakumulasikan

Provinsi	Rata-Rata	Provinsi	Rata-Rata
Aceh	10,24586	Nusa Tenggara Barat	95,52507
Sumatera Utara	185,7983	Nusa Tenggara Timur	12,49079
Sumatera Barat	36,21207	Kalimantan Barat	15,89014
R i a u	155,5093	Kalimantan Tengah	3,743429
J a m b i	2,625071	Kalimantan Selatan	7,595929
Sumatera Selatan	12,40857	Kalimantan Timur	35,5875
Bengkulu	0,43	Kalimantan Utara	1,595
Lampung	8,047714	Sulawesi Utara	16,42607
Kep Bangka Belitung	1,753571	Sulawesi Tengah	0,986
Kepulauan Riau	1008,307	Sulawesi Selatan	44,63714
DKI Jakarta	996,3017	Sulawesi Tenggara	1,72
Jawa Barat	218,9699	Gorontalo	0,685545
Jawa Tengah	76,51929	Sulawesi Barat	0,2105
DI Yogyakarta	132,7015	M a l u k u	8,197786
Jawa Timur	205,5488	Maluku Utara	0,429929
Banten	139,6588	Papua Barat	2,361909
B a l i	2803,507	P a p u a	12,105

Sumber: Badan Pusat Statistik

Setelah diakumulasikan dan dicari nilai rata-rata maka akan didapatkan nilai dari setiap variable. Kemudian data tersebut akan masuk ke tahapan *clustering* dengan menerapkan algoritma *K-means* untuk *mengcluster* data menjadi tiga cluster.

4.1. Centroid Data

Dalam penerapan algoritma *K-means* dihasilkan nilai titik tengah atau *centroid* dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 3. Penentuan titik cluster ini dilakukan dengan mengambil nilai terbesar (maksimum) untuk cluster

tinggi (C1), nilai rata-rata (average) untuk cluster sedang (C2) dan nilai terkecil (minimum) untuk cluster rendah (C3). Nilai titik tersebut dapat diketahui pada tabel berikut:

Tabel 2. *Centroid* Data Awal (Iterasi 1)

<i>Centroid</i>	
<i>Max (C1)</i>	2803,507
<i>Average (C2)</i>	183,963
<i>Min (C3)</i>	0,211

4.2. Clustering Data

Dengan menggunakan *centroid* tersebut maka dapat dicluster data yang telah didapat menjadi 3 cluster. Proses cluster dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data jumlah Tamu Asing pada Hotel Bintang Menurut Provinsi didapatkan pengelompokan pada iterasi 1 untuk 4 cluster tersebut. Cluster Tamu Asing tinggi (C1) yakni Bali. Cluster Tamu Asing sedang (C2) yakni 2 Provinsi dan cluster Tamu Asing rendah (C3) yakni 31 provinsi lainnya. Proses pencarian jarak terpendek, pengelompokan data pada iterasi berikut ini Clustering atas dapat digambarkan pada tabel dan gambar berikut:

Proses *K-means* akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya. Setelah mendapatkan nilai titik tengah atau *centroid*, proses sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terpendek, pengelompokan data pada iterasi terakhir dan Clustering data dapat digambarkan pada tabel berikut:

Tabel 3. *Centroid* Data Iterasi 4

<i>Centroid</i>	
<i>Max (C1)</i>	2803,507
<i>Average (C2)</i>	1002,304
<i>Min (C3)</i>	137,1008

Tabel 4. Perhitungan Jarak Pusat Cluster Iterasi 4

Provinsi	Rata-Rata	Iterasi 4			Jarak Terpendek
		c1	c2	c3	
Aceh	10,245857	2793,261	992,058	126,855	126,855
Sumatera Utara	185,79829	2617,709	816,506	48,698	48,698
Sumatera Barat	36,212071	2767,295	966,092	100,889	100,889
Riau	155,50929	2647,998	846,795	18,409	18,409
Jambi	2,6250714	2800,882	999,679	134,476	134,476
Sumatera Selatan	12,408571	2791,098	989,896	124,692	124,692
Bengkulu	0,43	2803,077	1001,87	136,671	136,671
Lampung	8,0477143	2795,459	994,257	129,053	129,053
Kep Bangka Belitung	1,7535714	2801,753	1000,55	135,347	135,347
Kepulauan Riau	1008,3067	1795,2	6,00251	871,206	6,003
DKI Jakarta	996,30171	1807,205	6,00251	859,201	6,003
Jawa Barat	218,96993	2584,537	783,334	81,869	81,869
Jawa Tengah	76,519286	2726,988	925,785	60,581	60,581
DI Yogyakarta	132,7015	2670,805	869,603	4,399	4,399
Jawa Timur	205,54879	2597,958	796,755	68,448	68,448
Banten	139,65879	2663,848	862,645	2,558	2,558
Bali	2803,5069	0	1801,2	2666,406	0,000
Nusa Tenggara Barat	95,525071	2707,982	906,779	41,576	41,576
Nusa Tenggara Timur	12,490786	2791,016	989,813	124,610	124,610
Kalimantan Barat	15,890143	2787,617	986,414	121,211	121,211
Kalimantan Tengah	3,7434286	2799,764	998,561	133,357	133,357
Kalimantan Selatan	7,5959286	2795,911	994,708	129,505	129,505

Provinsi	Rata-Rata	Iterasi 4			Jarak Terpendek
		c1	c2	c3	
Kalimantan Timur	35,5875	2767,919	966,717	101,513	101,513
Kalimantan Utara	1,595	2801,912	1000,71	135,506	135,506
Sulawesi Utara	16,426071	2787,081	985,878	120,675	120,675
Sulawesi Tengah	0,986	2802,521	1001,32	136,115	136,115
Sulawesi Selatan	44,637143	2758,87	957,667	92,464	92,464
Sulawesi Tenggara	1,72	2801,787	1000,58	135,381	135,381
Gorontalo	0,6855455	2802,821	1001,62	136,415	136,415
Sulawesi Barat	0,2105	2803,296	1002,09	136,890	136,890
Maluku	8,1977857	2795,309	994,106	128,903	128,903
Maluku Utara	0,4299286	2803,077	1001,87	136,671	136,671
Papua Barat	2,3619091	2801,145	999,942	134,739	134,739
Papua	12,105	2791,402	990,199	124,996	124,996

Tabel 5. Hasil Pengelompokan Iterasi 4

Provinsi	c1	c2	c3	Provinsi	c1	c2	c3
Aceh			1	Nusa Tenggara Barat			1
Sumatera Utara			1	Nusa Tenggara Timur			1
Sumatera Barat			1	Kalimantan Barat			1
Riau			1	Kalimantan Tengah			1
Jambi			1	Kalimantan Selatan			1
Sumatera Selatan			1	Kalimantan Timur			1
Bengkulu			1	Kalimantan Utara			1
Lampung			1	Sulawesi Utara			1
Kep Bangka Belitung			1	Sulawesi Tengah			1
Kepulauan Riau	1			Sulawesi Selatan			1
DKI Jakarta	1			Sulawesi Tenggara			1
Jawa Barat			1	Gorontalo			1
Jawa Tengah			1	Sulawesi Barat			1
DI Yogyakarta			1	Maluku			1
Jawa Timur			1	Maluku Utara			1
Banten			1	Papua Barat			1
Bali	1			Papua			1

4.3. Analisa Data

Pada iterasi 4, pengelompokan data yang dilakukan terhadap 4 cluster dengan iterasi 1 didapatkan hasil yang sama. Dari 34 data jumlah Tamu Asing berdasarkan provinsi dapat diketahui, 1 provinsi cluster tingkat tinggi 2 provinsi cluster tingkat sedang dan 31 provinsi lainnya termasuk cluster tingkat rendah

4. KESIMPULAN

Untuk melakukan penilaian Jumlah Tamu Asing menurut provinsi dapat menerapkan metode clustering *K-means*. Data diolah untuk memperoleh nilai dari Jumlah Tamu Asing. Data tersebut diolah menggunakan *Ms. Excel* untuk ditentukan nilai *centroid* dalam 3 cluster yaitu cluster tinggi (C1), cluster sedang (C2) dan cluster rendah (C3). Sehingga diperoleh penilaian berdasarkan pengelompokan Jumlah Tamu Asing berdasarkan provinsi dengan 1 provinsi dalam cluster C1, 2 provinsi cluster C2, dan 31 provinsi lainnya cluster C3. Hasil yang didapat dari penelitian bahwa hampir 90% provinsi di Indonesia masih memiliki potensi wisata yang rendah. Itu dibuktikan dari jumlah tamu asing yang menginap di hotel berdasarkan laporan dari badan pusat statistik. Hal ini menjadi masukan bagi pemerintah, untuk meningkatkan pertimbangan pemerintah khususnya Dinas Pariwisata dan Ekonomi Kreatif di Provinsi-provinsi untuk mengembangkan potensi objek daya tarik wisata di provinsi-provinsi serta sebagai bahan pertimbangan pengembangan hotel-hotel berbintang yang ada di Indonesia dan bertujuan untuk meningkatkan devisa negara.

REFERENCES

- [1] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means



- Clustering,” *Techno.COM*, vol. 16, no. 4, pp. 348–357, 2017.
- [2] T. Imandasari and A. P. Windarto, “Sistem Pendukung Keputusan dalam Merekomendasikan Unit Terbaik di PDAM Tirta Lihou Menggunakan Metode Promethee,” *J. Teknol. dan Sist. Komput.*, vol. 5, no. 4, p. 159, 2017.
- [3] alkhairi putrama, “Analisis dalam menentukan produk bri syariah terbaik berdasarkan dana pihak ketiga menggunakan ahp,” *putrama alkhairi*, vol. 3, no. 1, pp. 60–64, 2018.
- [4] V. Handayani, A. dan A. P. kurniati, “Analisa Clustering Menggunakan Algoritma K-Modes,” *Telkom University*, pp. 1-8, 2010.
- [5] N. Atthina dan L. Iswari, “Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat kesehatan Daerah dengan Metode *K-means*,” *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Vol. %1 dari %2ISSN 1907 - 5022, pp. B52 - B59, 2014.
- [6] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, pp. 26–33, 2017.
- [7] M. G. Sadewo, A. P. Windarto, and D. Hartama, “PENERAPAN DATAMINING PADA POPULASI DAGING AYAM RAS PEDAGING DI INDONESIA BERDASARKAN PROVINSI MENGGUNAKAN K-MEANS,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.
- [8] D. R. Sari, A. P. Windarto, D. Hartama, and S. Solikhun, “Sistem Pendukung Keputusan untuk Rekomendasi Kelulusan Sidang Skripsi Menggunakan Metode AHP-TOPSIS,” *J. Teknol. dan Sist. Komput.*, vol. 6, no. 1, p. 1, 2018.
- [9] Agus Perdana Windarto, “Implementasi Jst Dalam Menentukan Kelayakan Nasabah Pinjaman Kur Pada Bank Mandiri Mikro Serbelawan Dengan Metode Backpropagation,” *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 1, no. 1, pp. 12–23, 2017.
- [10] P. P. P. A. N. W. F. I. R. H. Zer and A. P. Windarto, “Analisis Pemilihan Rekomendasi Produk Terbaik Prudential Berdasarkan Jenis Asuransi Jiwa Berjangka Untuk Kecelakaan Menggunakan Metode Analytic Hierarchy Process (Ahp),” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 1, pp. 78–82, 2018