

SISTEM REKOMENDASI PROMOSI HOTEL PADA WISATAWAN MANCANEGERA BERBASIS DATA MINING

Anik Andriani

Manajemen Informatika AMIK BSI Yogyakarta
Jl Ring road Barat, Ambarketawang, Gamping, Sleman, Yogyakarta
Email : anik.andriani2011@gmail.com

Abstrak

Hotel merupakan salah satu dari akomodasi yang merupakan unsur penting dalam perkembangan dunia pariwisata. Promosi hotel ke mancanegara perlu dilakukan melihat banyaknya wisatawan mancanegara (wisman) yang datang ke Indonesia khususnya Yogyakarta. Namun kegiatan promosi yang tidak efektif berdampak pada penggunaan biaya promosi yang tidak tepat dan tidak sesuai pasar. Sistem rekomendasi promosi hotel ke mancanegara berbasis data mining dibuat dengan melakukan proses clustering pada data tamu hotel di Yogyakarta dari wisman. Clustering dilakukan dengan algoritma k-Means. Hasil dari proses clustering berupa cluster data tamu hotel dari wisman berdasarkan waktu kedatangan dan asal negara, sehingga diperoleh waktu kedatangan wisman tertinggi, dan asal negara wisman tertinggi. Hasil tersebut diimplementasikan dalam membangun program rekomendasi promosi hotel untuk menentukan waktu yang tepat untuk melakukan promosi dan negara yang tepat untuk melakukan promosi.

Kata kunci: Clustering, k-Means, Sistem rekomendasi promosi

1. Pendahuluan

1.1. Latar Belakang

Industri terbesar kedua di dunia setelah industri migas yang mampu menjadi primadona penghasil devisa negara yaitu industri pariwisata [1]. Indonesia sendiri merupakan negara yang menjadi salah satu tujuan pariwisata dunia. Propinsi Daerah Istimewa Yogyakarta (DIY) merupakan tempat kedua tujuan pariwisata di Indonesia setelah Bali. Hal ini menjadi alasan wisatawan mancanegara memilih Yogyakarta sebagai tempat tujuan pariwisata setelah mengunjungi Bali. Salah satu unsur di dalam industri pariwisata adalah akomodasi yang merupakan tempat bagi seseorang untuk tinggal sementara. Akomodasi dapat berupa hotel, losmen, *guest house*, pondok, *cottage*, *inn*, perkemahan, dan sebagainya. Dari berbagai macam akomodasi yang paling berpengaruh adalah hotel karena hotel merupakan salah satu akomodasi yang terbukti mempunyai jumlah kamar terbanyak dari semua jenis akomodasi [2].

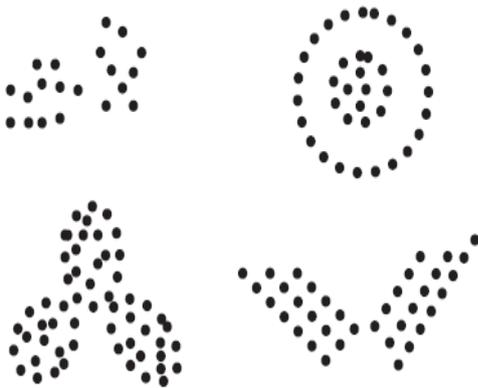
Promosi hotel ke mancanegara perlu dilakukan untuk meningkatkan jumlah tamu wisatawan mancanegara (wisman) yang menginap di hotel berbintang. Namun kegiatan-kegiatan promosi yang tidak efektif akan berpengaruh pada penggunaan biaya promosi yang tidak tepat dan tidak sesuai pasar. Karena itu perlu adanya strategi promosi hotel ke mancanegara yang tepat dan sesuai sasaran.

Penelitian ini bertujuan membuat sistem rekomendasi promosi hotel yang merekomendasikan waktu dan negara yang tepat untuk melakukan promosi sehingga lebih efektif dalam penggunaan biaya promosi. Sistem rekomendasi ini berbasis data mining yang menggunakan teknik *clustering* menggunakan algoritma *k-Means* dengan membuat segmentasi data tamu wisatawan mancanegara yang menginap di hotel berbintang di Yogyakarta. Hasil dari segmentasi data tamu hotel dari mancanegara diterapkan dalam pembuatan program untuk rekomendasi promosi hotel. Program yang dibuat menggunakan Microsoft Visual Basic 6.0.

1.2. Tinjauan Pustaka

Data mining sendiri merupakan sebuah proses untuk menemukan pola dari sebuah set data dimana pola tersebut harus otomatis atau semi otomatis dan pola yang dihasilkan harus memberi keuntungan [3]. *Clustering* merupakan pengelompokan bersama objek-objek yang mirip satu sama lain dan berbeda dengan objek *cluster* yang lain [4].

Clustering merupakan salah satu teknik dari data mining. *Clustering* adalah mencari kelompok (*cluster*) dari objek data berdasarkan persamaan data tersebut (kemiripan) sehingga masing-masing kelompok dari yang diperoleh mempunyai kesamaan yang besar, sementara kelompok yang lain cukup mempunyai perbedaan [5]. *Clustering* juga dapat diartikan sebagai pengelompokan data ke dalam kelas atau *cluster* sehingga objek di dalam *cluster* memiliki kesamaan yang tinggi satu sama lain dibandingkan dengan kelompok lainnya, tetapi sangat berbeda dengan objek dalam *cluster* lain [6]. Algoritma *clustering* mempartisi objek data seperti pola, entitas, kasus, *observances*, unit) menjadi sejumlah *cluster* (*group*, *subset*, atau kategori) seperti pada gambar 1 [7].

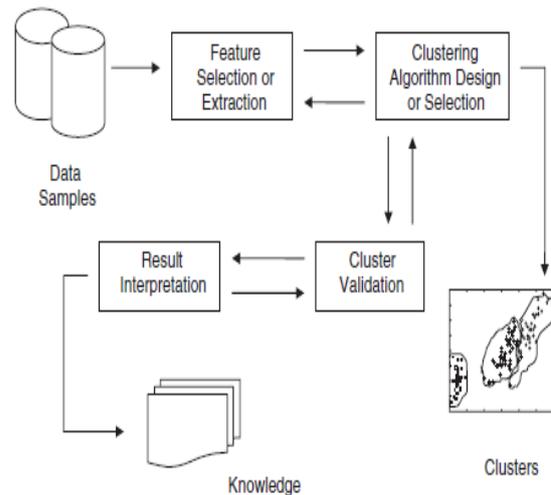


Gambar 1. Ilustrasi subjektivitas analisis cluster
Sumber [7]

Pada gambar 1 dapat dilihat ilustrasi subjektivitas analisis cluster yang menunjukkan sebuah cluster dapat terbentuk karena adanya kedekatan antara data satu dengan yang lainnya. Bentuk cluster yang terbentuk dapat berbeda-beda tergantung dari kedekatan data yang ada.

Analisis cluster merupakan sebuah aktivitas yang penting karena dengan melakukan clustering otomatis dapat membuat identifikasi daerah yang padat dan jarang dalam sebuah ruang obyek dan dapat menemukan pola distribusi secara keseluruhan serta korelasi yang menarik antar atribut data. Analisis cluster telah banyak digunakan dalam berbagai aplikasi seperti riset pasar, pengenalan pola, analisis data, dan pengolahan gambar. Clustering dalam dunia bisnis digunakan untuk membantu pelaku marketing dalam menemukan kelompok pelanggan berdasarkan pola pembelian. Dalam bidang biologi clustering dapat digunakan untuk memperoleh taksonomi tumbuhan dan hewan dengan cara mengkategorikan gen dengan fungsi serupa dan mendapatkan pengetahuan ke dalam struktur yang melekat dalam populasi. Dalam bidang ini juga dapat digunakan untuk membangun mengidentifikasi daerah pengguna lahan yang sama dalam sebuah database dan mengidentifikasi kelompok rumah di kota sesuai dengan tipe rumah, nilai, dan lokasi geografis, serta identifikasi kelompok pemegang polis asuransi mobil dengan biaya klaim rata-rata tinggi. Dalam bidang teknologi informasi clustering dapat digunakan untuk mengklasifikasi dokumen web untuk penemuan informasi [6].

Clustering dapat dilakukan dengan beberapa algoritma antara lain *k-Means*, SOM, EM, BIRCH, atau algoritma clustering lainnya. Beberapa faktor yang perlu dipertimbangkan dalam memilih metode atau algoritma yang akan digunakan dalam clustering yaitu keahlian teknis yang tersedia (*hardware* dan *software*), dan kinerja metode tersebut (Shi et al, 2011). Prosedur analisis cluster dapat dilihat pada gambar 2.



Gambar 2. Prosedur clustering
Sumber [7]

Gambar 2 menggambarkan prosedur analisis cluster dengan empat langkah dasar sebagai berikut [7]:

- 1) *Feature selection or extraction*, fitur seleksi merupakan tahap memilih fitur yang membedakan dari satu set calon, sedangkan ekstraksi fitur menggunakan beberapa transformasi untuk menghasilkan fitur yang berguna.
- 2) *Clustering algorithm design or selection*, tahap ini digunakan untuk menentukan ukuran yang tepat dan membangun kedekatan fungsi kriteria. Dalam hal ini objek data dikelompokkan ke dalam kelompok yang berbeda sesuai dengan kemiripan data satu sama lain.
- 3) *Cluster validation*, pada tahap ini dilakukan penilaian dan pengujian hasil pengelompokan.
- 4) *Result interpretation*, tahap ini merupakan tahap analisis dan eksperimen lebih lanjut dari hasil pengelompokan data.

Algoritma *k-Means* merupakan metode paling terkenal dalam clustering. Metode ini merupakan salah satu metode pengelompokan data yang lebih efisien dan efektif dalam menentukan kelompok bentuk bulat [8]. Algoritma *k-Means* mengambil parameter input, *k*, dan partisi satu set objek ke dalam cluster *k* sehingga kesamaan intra cluster tinggi tapi kesamaan inter cluster rendah. Kesamaan cluster yang diukur dalam kaitannya dengan nilai rata-rata dari objek di dalam cluster yang dapat dilihat sebagai sebuah cluster's centroid atau pusat gravitasi [6].

Langkah-langkah membuat clustering dengan algoritma *k-Means* antara lain [4]:

- 1) Menentukan jumlah cluster *k* yang akan dipakai. Nilai *k* pada umumnya merupakan integer kecil seperti 2,3,4, atau 5, tapi bisa lebih besar.
- 2) Menentukan pusat cluster (*centroid*) secara random/acak. Pusat cluster merupakan nilai rata-rata (*mean*) dari semua data atau objek dalam sebuah cluster.

- 3) Tempatkan setiap objek data pada *cluster* yang mempunyai nilai *centroid* terdekat. Kedekatan kedua objek ditentukan berdasarkan jarak yang dipakai pada algoritma *k-Means* yaitu *Euclidean distance* (d).

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$x = x_1, x_2, \dots, x_n$, dan $y = y_1, y_2, \dots, y_n$ merupakan banyaknya n atribut (kolom) antar dua objek data.

- 4) Menghitung ulang nilai *centroid* dari *cluster* k yang baru.
 5) Ulangi langkah 3 dan 4 sampai *centroid* tidak lagi bergerak/stabil dan tidak ada penurunan nilai yang signifikan dari nilai SSE (*Sum of Squared Errors*).

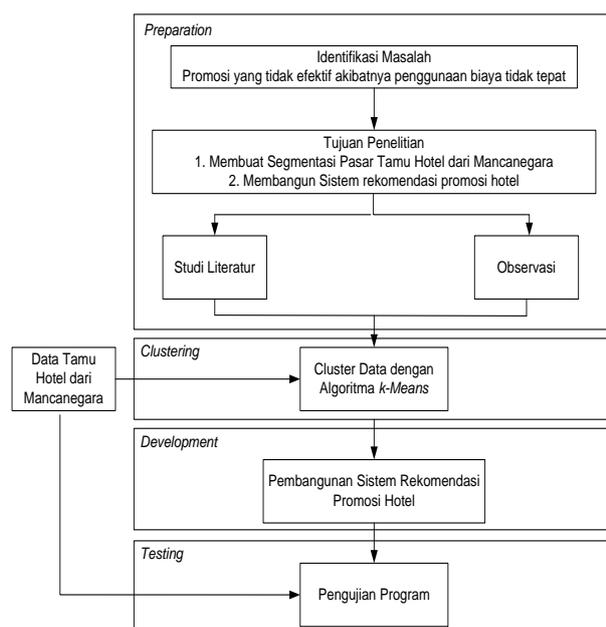
$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2 \quad (2)$$

$p \in C_i$ = tiap data poin pada cluster i , m_i = *centroid* dari cluster i , d = jarak/ *distances/ variance* terdekat pada masing-masing cluster i

1.3. Metode Penelitian

Penelitian ini menggunakan algoritma *k-Means* untuk membuat *clustering* data tamu wisman yang menginap di hotel di wilayah Yogyakarta. Populasi dari penelitian ini adalah data tamu wisman yang menginap di hotel di wilayah Yogyakarta yang didapat dari Buku Statistik Pariwisata Yogyakarta. Sampel yang digunakan adalah data tahun 2010 sampai dengan tahun 2012 yang menggunakan jasa akomodasi sejumlah 36 negara.

Tahap-tahap yang akan dilakukan dalam membangun sistem rekomendasi promosi hotel ini digambarkan dalam kerangka pemikiran seperti pada gambar 1.



Gambar 1. Kerangka Pemikiran

Pada gambar 1 dapat dilihat penelitian ini terdiri dari empat tahap yaitu *preparation*, *clustering*, *development*, dan *testing*. Tahap *preparation* merupakan tahap yang dilakukan untuk mengidentifikasi masalah, merumuskan tujuan penelitian, dan tahap literatur dan observasi. Tahap yang kedua yaitu tahap *clustering* yang merupakan tahap membuat *cluster* data dengan menggunakan algoritma *K-Means*. Data yang digunakan untuk *clustering* diambil dari buku Statistik Pariwisata Yogyakarta tahun 2010 sampai dengan 2012 yang berisi data tamu hotel dari mancanegara di Yogyakarta. Hasil dari proses *clustering* berupa *clustering* data berdasarkan waktu kedatangan dan asal negara yang selanjutnya diterapkan dalam tahap *development* yaitu tahap pembangunan sistem rekomendasi promosi hotel. Tahap terakhir adalah *testing* yang bertujuan untuk melakukan pengujian pada sistem yang telah dibangun.

2. Pembahasan

2.1. Proses Clustering

Hasil dari proses *clustering* data wisatawan mancanegara yang menginap di hotel berbintang di propinsi DIY berdasarkan waktu menginapnya dapat dilihat pada tabel 1.

Tabel 1. *Cluster* data tamu wisman di hotel berbintang DIY berdasarkan waktu kedatangan

Cluster ke-	Bulan
1	Juli
2	Agustus, September
3	Mei, Juni, Oktober
4	Januari, Februari, Maret, April
5	November, Desember

Pada tabel 1 menunjukkan *cluster* data tamu wisman di hotel berbintang di DIY berdasarkan waktu kedatangan. Dapat dilihat waktu kedatangan wisman yang menginap di hotel tertinggi ada di bulan Juli, diikuti dengan bulan Agustus dan September. Bulan Mei, Juni, dan Oktober jumlah tamu yang menginap di hotel masuk kategori sedang. Sedangkan jumlah rendah ada di bulan Januari, Februari, Maret, April, November, Desember.

Hasil dari proses *clustering* data wisman yang menginap di hotel berbintang di propinsi DIY berdasarkan asal negara wisman dapat dilihat pada tabel 2.

Tabel 2. *Cluster* data tamu wisman di hotel berbintang DIY berdasarkan asal negara

Cluster ke-	Asal Negara
1	Belanda
2	Jepang, Malaysia, Perancis
3	Amerika Serikat, Australia, Jerman,

	Slam/Muangthai
4	Belgia, Eropa Lainnya, India, Inggris, Italia, Korea Selatan, RRC, Spanyol/Portugal, Swiss
5	Amerika Latin, ASEAN Lainnya, Asia Lainnya, Asia Pasifik Lainnya, Austria, Brunei Darussalam, Canada, Denmark, Finlandia, Hongkong, Afrika, Norwegia, Philipina, Rusia, Selandia Baru, Singapura, Swedia/Skandinavia, Taiwan, Timur Tengah

Tabel 2 menunjukkan *cluster* data tamu dari wisman yang menginap di hotel di DIY. Dari tabel tersebut dapat dilihat asal negara wisman terbanyak berasal dari Belanda, yang diikuti dengan Jepang, Malaysia, dan Perancis. Amerika Serikat, Australia, Jerman, dan Slam/Muangthai jumlah wisman yang menginap masuk kategori sedang. Selain itu asal negara wisman yang lain masuk kategori rendah.

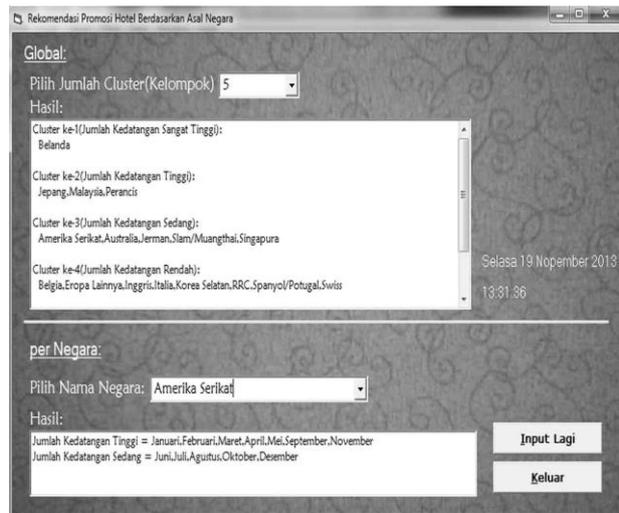
2.2. Pembangunan Sistem

Hasil dari *clustering* diimplementasikan ke dalam pembuatan program untuk rekomendasi promosi hotel ke mancanegara. Program yang dibuat terdiri dari menu utama seperti pada gambar 2. Dari gambar 2 dapat dilihat menu utama terdapat dua pilihan menu yaitu menu promosi berdasarkan asal negara untuk melihat hasil rekomendasi negara yang tepat untuk melakukan promosi. Menu kedua yaitu menu promosi berdasarkan bulan untuk melihat hasil rekomendasi bulan yang tepat untuk melakukan promosi di berbagai negara.



Gambar 2. Menu Utama

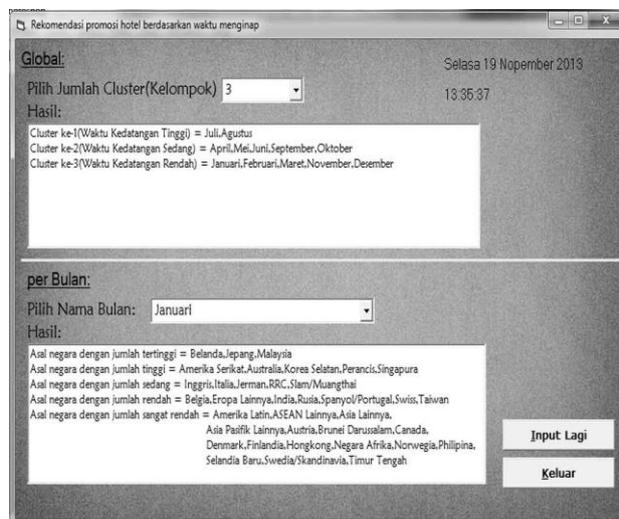
Tampilan menu rekomendasi promosi hotel berdasarkan asal negara dapat dilihat pada gambar 3.



Gambar 3. Menu rekomendasi berdasarkan asal negara

Form menu rekomendasi promosi hotel berdasarkan asal negara wisman menyediakan pilihan untuk melihat cluster data secara global maupun per negara. Dari pilihan rekomendasi secara global dapat dipilih jumlah cluster yang diinginkan. Dari rekomendasi per negara dapat dilihat waktu menginap tertinggi maupun terendah berdasarkan negara yang dipilih.

Menu rekomendasi promosi hotel berdasarkan waktu menginap wisman dapat dilihat pada gambar 4.



Gambar 4. Menu rekomendasi berdasarkan waktu menginap

Form menu rekomendasi promosi hotel berdasarkan waktu wisman menginap menyediakan pilihan untuk melihat cluster data secara global maupun berdasarkan nama bulan. Dari pilihan rekomendasi secara global dapat dipilih jumlah *cluster* yang diinginkan. Dari rekomendasi per bulan dapat dilihat lebih rinci berdasarkan bulan yang dipilih.

2.3. Pengujian Sistem

Pengujian sistem digunakan untuk menguji program rekomendasi promosi hotel yang sudah dibuat. Pengujian program menggunakan metode pengujian *black box*. Pengujian *black box* terfokus pada fungsionalitas dari perangkat lunak. *Black box testing* fokus menemukan fungsi yang tidak benar, kesalahan antarmuka, kesalahan pada struktur data, kesalahan performansi, dan kesalahan inisialisasi.

Hasil dari pengujian program dengan metode *black box* dapat dilihat pada tabel 3. Pengujian menggunakan data training yang digunakan untuk membuat cluster dan data testing (data yang baru) yang diujikan pada hasil cluster data berdasarkan asal negara dan cluster data berdasarkan waktu menginap.

Tabel 3. Hasil pengujian dengan metode *black box*

Jenis Pengujian	Tingkat performansi	
	Data Training	Data Testing (Data Baru)
Pengujian cluster data berdasarkan asal negara	89%	82%
pengujian cluster data berdasarkan waktu menginap	75%	75%

3. Kesimpulan

Hasil *clustering* data tamu yang menginap di hotel berbintang di Yogyakarta yang berasal dari mancanegara dengan algoritma k-Means menghasilkan *cluster* data berdasarkan asal negara dan waktu menginap. Hasil dapat digunakan untuk rekomendasi promosi hotel ke mancanegara dalam menentukan waktu yang tepat untuk promosi dan negara-negara yang tepat untuk promosi.

Belanda, Jepang, Malaysia, dan Perancis merupakan negara-negara asal wisman tertinggi yang menginap di hotel yang ada di kawasan Yogyakarta, sedangkan untuk waktu bulan Juli, Agustus, September merupakan waktu dengan tingkat tertinggi untuk jumlah wisman yang menginap di hotel berbintang di Yogyakarta sehingga dapat direkomendasikan melakukan promosi ke mancanegara pada bulan-bulan sebelumnya. Hasil selengkapnya dari *clustering* data dapat dilihat pada program rekomendasi promosi hotel ke mancanegara yang sudah dibuat.

Hasil pengujian di program rekomendasi promosi hotel ke mancanegara dengan metode *black box* dimana data yang digunakan untuk pengujian yaitu data *training* dan

data *testing* (dengan data baru) menunjukkan tingkat performansi program yang tinggi.

Daftar Pustaka

- [1] I. G. Pitana and P. G. Gayatri, *Sosiologi Pariwisata*. Yogyakarta: Andi, 2005.
- [2] E. Sugiarto and S. Sulartiningrum, *Pengantar Akomodasi dan Restoran*. Jakarta: Gramedia, 2003.
- [3] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques 3rd Edition*. Burlington: Elsevier, 2011.
- [4] M. Bramer, *Principles of Data Mining*. London: Springer, 2007.
- [5] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Berlin: Springer, 2011.
- [6] J. Han and M. Kamber, *Data Mining Concepts And Techniques 2nd Edition*. San Fransisco: Elsevier, 2006.
- [7] R. Xu and D. C. Wunsch II, *Clustering*. New Jersey: John Willey and Sons, 2009.
- [8] C. Vercellis, *Business Intelligence Data Mining and Optimization for Decision Making*. Cornwall: John Willey and Son, 2009.

Biodata Penulis

Anik Andriani, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK Nusa Mandiri Jakarta, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2012. Saat ini menjadi dosen di AMIK BSI Yogyakarta.