



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

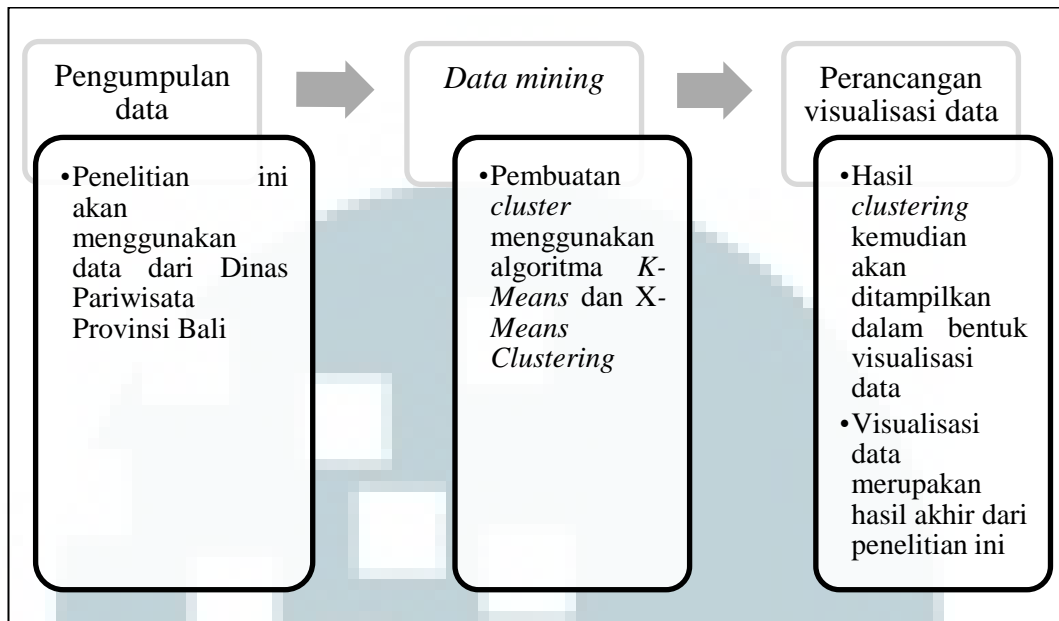
METODOLOGI PENELITIAN

3.1. Objek Penelitian

Jumlah wisatawan domestik yang berkunjung ke Bali meningkat sebanyak 11,77% dari tahun 2014 ke 2015 (Disprada, 2017), juga pertumbuhan kunjungan wisatawan mancanegara di Bali meningkat 17,85% dari tahun 2016 ke 2017 (Kemenpar, 2017). Hal ini memperlihatkan bahwa provinsi Bali masih menjadi destinasi yang diminati oleh para wisatawan, baik wisatawan domestik maupun mancanegara, oleh karena hal itu maka dipilihlah objek pariwisata di provinsi Bali sebagai objek dalam penelitian ini.

3.2. Metode Penelitian

Dalam penelitian ini digunakan algoritma *K-Means* dan *X-Means Clustering* untuk mengelompokkan data objek pariwisata di provinsi Bali menjadi beberapa *cluster* berdasarkan jumlah pengunjungnya. Penelitian ini terdiri dari 3 tahap utama, yang dapat dilihat pada gambar 3.1.



Gambar 3.1. Tahapan Utama Penelitian

Penjelasan:

1. Pengumpulan data

Data-data yang diperlukan antara lain:

- a. Daftar objek pariwisata di provinsi Bali
- b. Jumlah pengunjung dalam setiap objek pariwisata di provinsi Bali
- c. Objek pariwisata di provinsi Bali berdasarkan daerahnya
- d. Objek pariwisata di provinsi Bali berdasarkan tipe objek pariwisatanya

Pengumpulan data akan dilakukan dengan mendatangi kantor Dinas Pariwisata Provinsi Bali.

2. *Data mining*

Data yang telah dikumpulkan akan digunakan dalam proses *clustering*. Proses *clustering* akan dilakukan dengan menggunakan algoritma *K-Means* dan *X-Means Clustering*.

3. Perancangan visualisasi data

Hasil dari proses *clustering* tersebut akan ditampilkan dalam bentuk visualisasi data yang disajikan di dalam sebuah *dashboard*.

3.2.1. Metode Penyelesaian Masalah

3.2.1.1. Metode Penentuan Jumlah *Cluster* Optimal

Ada dua metode yang dapat digunakan untuk penentuan jumlah *cluster* optimal, yang dapat dilihat pada tabel 3.1.

Tabel 3.1. Perbandingan antara *Bayesian Information Criterion (BIC)* dan *Davies-Bouldin Index (DBI)*

<i>Bayesian Information Criterion (BIC)</i> (Hennig, Meila, & Rocci, 2015)	<i>Davies-Bouldin Index (DBI)</i> (S. C & G., 2012)
Dalam kondisi tertentu, BIC konsisten dalam memilih jumlah <i>cluster</i> yang benar	Berdasarkan perhitungan nilai <i>Sum of Square within Cluster (SSW)</i> , <i>Sum of Square between Cluster (SSB)</i> , nilai rasio, dan <i>Davies-Bouldin Index</i>
Cenderung mendukung model dengan komponen yang lengkap untuk memberikan hasil yang baik	Nilai <i>cluster</i> optimum (<i>k</i>) ditentukan dari nilai DBI yang diperoleh yang paling mendekati 0

Dalam penelitian ini *Bayesian Information Criterion* diimplementasikan untuk perhitungan metode *X-Means* dalam RapidMiner dengan menggunakan operator *X-Means*. Dalam operator *X-Means* ini, sudah tertanam perhitungan *Bayesian Information Criterion*, sehingga jumlah *cluster* optimal dapat otomatis terhitung dan ditampilkan.

Sedangkan *Davies-Bouldin Index* diimplementasikan untuk perhitungan metode *K-Means* dalam RapidMiner dengan menggunakan operator *Cluster Distance Performance* yang hasilnya kemudian ditampilkan dalam operator *Log*, sehingga dapat dilihat nilai *Davies-Bouldin Index* untuk setiap jumlah *cluster*.

3.2.1.2. Metode Clustering

Ada dua metode *clustering* yang dapat digunakan dalam penelitian ini, yang dapat dilihat pada tabel 3.2.

Tabel 3.2. Perbandingan antara Algoritma *X-Means Clustering* dan Algoritma *K-Means Clustering*

Algoritma <i>X-Means Clustering</i> (Nguyen, Kowalczyk, Orłowski, & Ziółkowski, 2016).	Algoritma <i>K-Means Clustering</i>
Merupakan perluasan dari algoritma <i>K-Means Clustering</i>	Suatu objek hanya akan menjadi anggota satu <i>cluster</i> . (Cebeci & Yildiz, 2015)
Dapat mengidentifikasi jumlah <i>cluster</i> terbaik dengan sendirinya menggunakan <i>Bayesian Information Criterion</i> (BIC)	Menggunakan pendekatan yang bersifat <i>hit-and-miss</i> untuk menentukan jumlah <i>cluster</i> yang benar (Satapathy, Mandal, Udgata, & Bhateja, 2016)

Dalam penelitian ini algoritma *X-Means Clustering* dan *K-Means Clustering* ini digunakan untuk mengelompokkan data pada tahap *data mining* dan akan diimplementasikan menggunakan RapidMiner. Algoritma *X-Means Clustering* diimplementasikan dengan menggunakan operator *X-Means*. Dengan operator *X-Means*, dapat secara otomatis melakukan *clustering* menggunakan metode *X-Means Clustering* untuk data objek pariwisata Bali, dan kemudian dapat menampilkan jumlah *cluster* terbaik beserta anggota setiap *cluster*.

Sedangkan, algoritma *K-Means Clustering* diimplementasikan dengan menggunakan operator *K-Means*. Dengan operator *K-Means*, dapat secara otomatis melakukan *clustering* menggunakan metode *K-Means* untuk data objek pariwisata Bali, tetapi sebelumnya perlu ditentukan jumlah *cluster* optimum yang telah dihitung menggunakan perhitungan *Davies-Bouldin Index*.

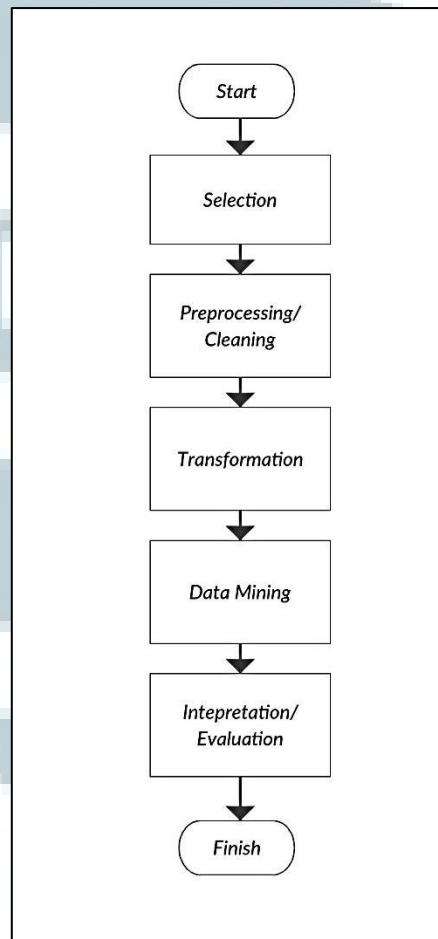
3.2.2. Metode Perancangan Visualisasi Data

Ada dua metode untuk membuat perancangan visualisasi data, yang dapat dilihat pada tabel 3.3.

Tabel 3.3. Perbandingan antara KDD dengan CRISP-DM

<i>Knowledge Discovery in Databases</i> (Shafique & Qaiser, 2014)	<i>Cross-Industry Standard Process for Data Mining</i> (Shafique & Qaiser, 2014)
Dapat digunakan untuk penelitian dan proyek yang sifatnya lebih umum	Bersifat “company oriented”, digunakan untuk proses bisnis
Lebih lengkap dan akurat	Ruang lingkup kurang lengkap

Karena penelitian ini tidak ditujukan untuk proses bisnis, maka dipilihlah metode KDD untuk digunakan dalam penelitian ini. Gambar 3.2. merupakan *flowchart* tahapan dalam KDD:



Gambar 3.2. Flowchart Tahap dalam Knowledge Discovery in Database (KDD)

Penjelasan:

1. *Selection*

Merupakan proses untuk pengumpulan dan pemilihan data. Data-data yang diperlukan antara lain:

- a. Daftar objek pariwisata di provinsi Bali
- b. Jumlah pengunjung dalam setiap objek pariwisata di provinsi Bali
- c. Objek pariwisata di provinsi Bali berdasarkan daerahnya
- d. Objek pariwisata di provinsi Bali berdasarkan tipe objek pariwisata

Pengumpulan data akan dilakukan dengan mendatangi kantor Dinas Pariwisata Provinsi Bali.

2. *Preprocessing/Cleaning*

Data yang sudah dikumpulkan akan dirapihkan dan “dibersihkan”.

Langkah-langkah yang dilakukan pada tahap ini yaitu:

- a. Menghapus baris dan kolom yang tidak digunakan
- b. Menyamakan penulisan nama objek pariwisata
- c. Menggabungkan setiap total pengunjung per tahun ke dalam 1 tabel yaitu tabel Total Pengunjung
- d. Membuat tabel Daftar Objek yang berisi nama objek pariwisata beserta tipe objek dan kabupaten/kota.

3. *Transformation*

Data yang sudah dirapihkan dan “dibersihkan” selanjutnya akan ditransformasi untuk memudahkan pada proses *data mining*. Dalam

penelitian ini, transformasi yang dimaksudkan yaitu mengubah data dengan format (tipe data) yang sesuai.

4. *Data Mining*

Data yang telah ditransformasikan akan dibuat menjadi beberapa kelompok berdasarkan jumlah pengunjungnya menggunakan metode *clustering* dengan algoritma *K-Means* dan *X-Means Clustering*.

5. *Intepretation/Evaluation*

Hasil *clustering* yang dilakukan pada tahap *data mining* dibuat visualisasi datanya agar informasi yang ditampilkan dapat dengan mudah ditangkap.

3.3. Teknik Pengumpulan Data

Data yang akan digunakan di dalam penelitian ini didapatkan secara langsung dengan mendatangi kantor Dinas Pariwisata Provinsi Bali di Denpasar, Bali. Data yang didapatkan yaitu data kunjungan wisatawan nusantara dan mancanegara dari tahun 2012 hingga 2016.

3.4. Variabel Penelitian

3.3.1. Variabel Independen (Variabel X)

Dalam penelitian ini, yang menjadi variabel independen (variabel X) dalam penelitian ini adalah:

X_1 = Jumlah pengunjung tahun 2012

X_2 = Jumlah pengunjung tahun 2013

X_3 = Jumlah pengunjung tahun 2014

X_4 = Jumlah pengunjung tahun 2015

X_5 = Jumlah pengunjung tahun 2016

3.3.2. Variabel Dependen (Variabel Y)

Dalam penelitian ini, yang menjadi variabel dependen (variabel Y) dalam penelitian ini adalah:

Y = Objek pariwisata yang telah dikelompokkan

3.5. Teknik Pengolahan Data

3.4.1. Tools untuk Data Mining

Terdapat 2 *tools* yang dapat digunakan untuk proses *data mining*, yang dapat dilihat pada tabel 3.4.

Tabel 3.4. Perbandingan antara RapidMiner dengan MATLAB

RAPIDMINER (Hofmann & Klinkenberg, 2016)	MATLAB
<i>Open source</i>	Berbayar (Chapman, 2015)
Menyediakan sebuah <i>integrated environment</i> untuk semua langkah proses <i>data mining</i>	Memiliki banyak built-in tools untuk memecahkan masalah dan mengembangkan ilustrasi grafis (Attaway, 2016)
<i>Graphical User Interface</i> (GUI) yang mudah digunakan untuk perancangan proses <i>data mining</i> , visualisasi data dan hasil, validasi dan optimalisasi proses, dan untuk <i>automated deployment</i> serta integrasi ke sistem yang lebih	Memiliki built-in functions untuk melakukan banyak operasi (Attaway, 2016)

RAPIDMINER (Hofmann & Klinkenberg, 2016)	MATLAB
kompleks	
Memungkinkan untuk menyebarkan hasil ke semua jenis data termasuk <i>file, databases, time series data</i> , data teks, halaman web, media sosial, data gambar, data audio, <i>web services</i> , dan banyak sumber data lainnya	Sebuah interpreted language dan proses eksekusi lebih lambat dari tools lainnya (Chapman, 2015)

Karena memiliki *step-step* yang lengkap untuk proses *data mining*, GUI yang baik, serta dapat menyebarkan ke dalam berbagai macam sumber data, maka dipilihlah RapidMiner untuk digunakan dalam tahap *data mining*.

3.4.2. Tools untuk Data Preprocessing dan Visualisasi Data

Terdapat 2 *tools* yang dapat digunakan untuk tahap *data preprocessing* serta pembuatan visualisasi data, yang dapat dilihat pada tabel 3.5.

Tabel 3.5. Perbandingan antara Microsoft Power BI dengan Tableau

MICROSOFT POWER BI (Aspin, 2016)	TABLEAU (Milligan, 2016)
Dapat menemukan, membersihkan data, dan mengumpulkan data sehingga dapat mengembangkan dan hasilnya dapat disampaikan dalam bentuk visualisasi data	Dapat <i>connect</i> ke berbagai sumber data dan mengintegrasikan data
Dapat membuat <i>data model</i>	Dapat melakukan explorasi dan pembuatan visualisasi data
Dapat membagikan <i>reports</i> dengan <i>audience</i> yang ditunjuk melalui Azure, dan dapat melihat <i>insights</i> menggunakan Windows, iOS, dan Android <i>device</i>	Dapat share hasil visualisasi dan <i>dashboard</i> dengan pengguna Tableau Desktop dan Tableau Reader

Karena Microsoft Power BI dapat membuat *data model* dan hasil *report* dapat di-*share* ke berbagai *device*, maka dipilihlah *Microsoft Power BI* untuk digunakan dalam proses *preprocessing data* serta visualisasi data.

3.6. Penelitian Terdahulu

Penyusunan penelitian ini mengambil beberapa referensi penelitian sebelumnya termasuk jurnal-jurnal yang berhubungan dengan penelitian ini yang dituliskan dalam tabel 3.6.

Tabel 3.6. Penelitian Terdahulu

Judul Jurnal	Pembahasan	Kesimpulan
<p>ANALISIS OBJEK DAYA TARIK WISATA FAVORIT BERDASARKAN JUMLAH PENGUNJUNG DI DAERAH ISTIMEWA YOGYAKARTA</p> <p>Peneliti: Atun Yulianto</p> <p>Lokasi: Akademi Pariwisata BSI Yogyakarta</p> <p>Nama Jurnal: Jurnal Media Wisata, Volume 2015, Nomor 2, November 2017</p> <p>Halaman: 555-567</p>	<p>Penelitian ini bertujuan untuk mengetahui objek wisata di Daerah Istimewa Yogyakarta yang merupakan favorit wisatawan berdasarkan jumlah pengunjung untuk tahun 2011 hingga 2015.</p> <p>Metode penelitian yang digunakan adalah deskriptif kualitatif yang didukung data kuantitatif untuk memberikan gambaran matematis dari tiga objek wisata yang paling dominan dikunjungi oleh wisatawan.</p> <p>Metode pengumpulan data yang dilakukan meliputi studi pustaka, dokumentasi, dan memanfaatkan teknologi informasi internet sebagai sarana untuk memperoleh informasi</p>	<p>Objek wisata yang banyak dikunjungi dan tidak banyak dikunjungi oleh wisatawan sebagai berikut:</p> <ol style="list-style-type: none"> 1. Pantai Parangtritis Kabupaten Bantul menjadi favorit wisatawan nusantara dan domestik disusul Candi Prambanan Kabupaten Sleman dan Gembira Loka Kota Yogyakarta. 2. Museum, candi-candi kecil, dan makam-makam bersejarah menjadi objek daya tarik wisata yang memiliki jumlah pengunjung wisatawan yang masih terbatas.

Judul Jurnal	Pembahasan	Kesimpulan
	kuantitatif dari sumber-sumber kepariwisataan di DIY.	

