



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 tamu di hotel memang bukan indikator utama dalam melihat perkembangan pariwisata. Di lokasi tertentu, penghuni hotel bukanlah wisatawan, melainkan pebisnis, namun perhotelan menjadi salah satu penopang utama pariwisata (Islahuddin, 2017). Bila dilihat sepanjang 12 tahun terakhir (2004-2016), kenaikan secara konsisten terjadi pada Tingkat Penghunian Kamar Hotel (TPH) untuk hotel berbintang (Beritagar, 2017). Kenaikan itu seiring dengan jumlah kamar yang tersedia dalam 12 tahun terakhir. Jumlah kamar untuk hotel berbintang naik dari 170 ribu pada 2004, menjadi 294 ribu pada 2016 (BPS, 2017). Hal ini memperlihatkan bahwa jumlah tamu yang datang ke hotel berbintang terjadi kenaikan secara terus menerus, oleh karena hal tersebut maka dipilihlah jumlah tamu per hari di seluruh Indonesia sebagai objek dalam penelitian ini.

3.2. Metode Penelitian

Dalam penelitian ini digunakan algoritma *Multiple Linear Regression* untuk menemukan rumus agar dapat memprediksi data jumlah tamu per hari di seluruh Indonesia dan membuat visualisasi mengenai hasil yang telah ditemukan.

Tabel 3.1. menjadi pedoman algoritma yang akan dipakai pada penelitian ini.

Tabel 3.1. Tabel Perbandingan Algoritma

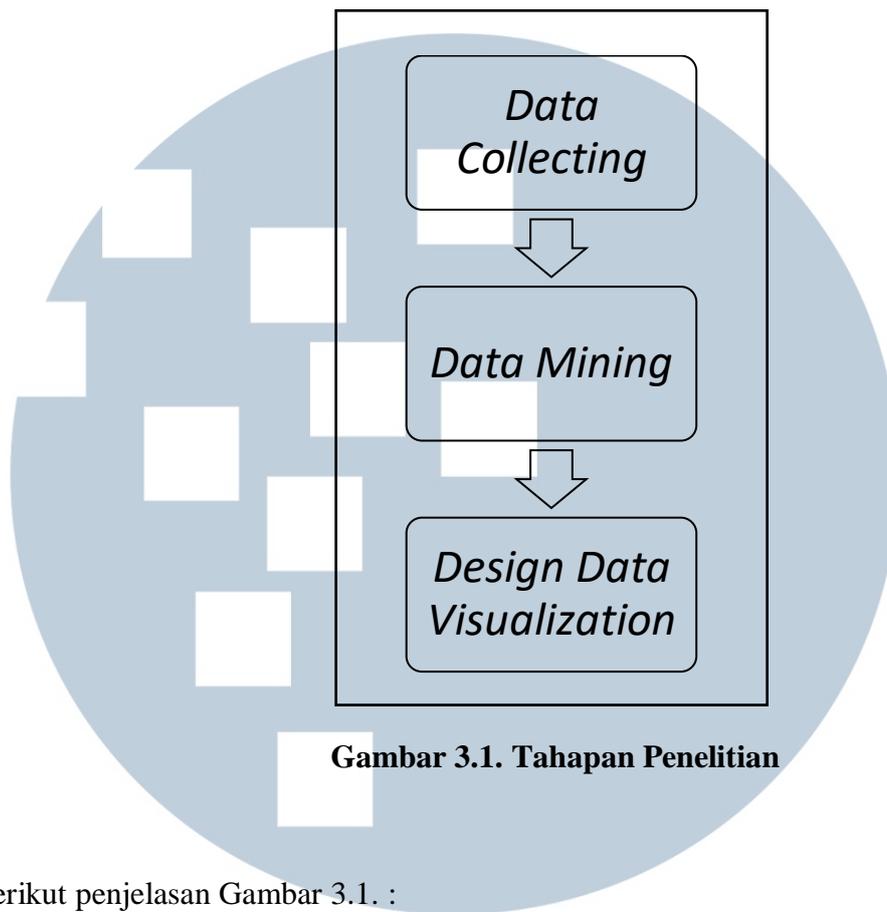
<i>Multiple Linear Regression</i> (Kelvin, 2018)	<i>Decision Tree</i> (Kelvin, 2018)
Menangani masalah regresi	Menangani masalah regresi dan klasifikasi
Mudah untuk menjelaskan hasil prediksi ke orang lain	Sedikit susah menjelaskan hasil prediksi ke orang lain
Selisih akurasi prediksi rendah	Selisih akurasi prediksi rendah
Kecepatan <i>Training</i> cepat	Kecepatan <i>Training</i> cepat
Kecepatan prediksi cepat	Kecepatan prediksi cepat

Dalam penelitian ini *Multiple Linear Regression* digunakan karena metode ini berguna untuk menangani masalah regresi dan mudah untuk menjelaskan hasil prediksi ke orang lain.

Multiple Linear Regression biasanya berguna untuk menganalisis suatu hubungan sebab akibat antara variabel dependen dan variabel independen. *Multiple Linear Regression* digunakan untuk membuat suatu rumus regresi yang berguna untuk meramalkan suatu variabel.

Penelitian ini terdiri dari 3 tahapan agar dapat sampai menjadi visualisasi data, yaitu: *data collecting*, *data mining*, *design data visualization*.

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.1. Tahapan Penelitian

Berikut penjelasan Gambar 3.1. :

1. *Data Collecting*

Data- data yang diperlukan untuk dapat memprediksi antara lain:

- a. Jumlah akomodasi pada hotel berbintang,
- b. Jumlah pekerja pada hotel berbintang,
- c. Jumlah tamu pada hotel berbintang.

Data akan didapatkan dengan cara mengunduh *file* dari *website* BPS.

2. *Data Mining*

Data yang telah di-*collect* akan digunakan dalam proses prediksi. Proses prediksi akan dilakukan dengan menggunakan algoritma *Multiple Linear Regression*.

3. *Design Data Visualization*

Hasil yang didapatkan dari proses prediksi akan ditampilkan dalam bentuk *dashboard* di PowerBI.

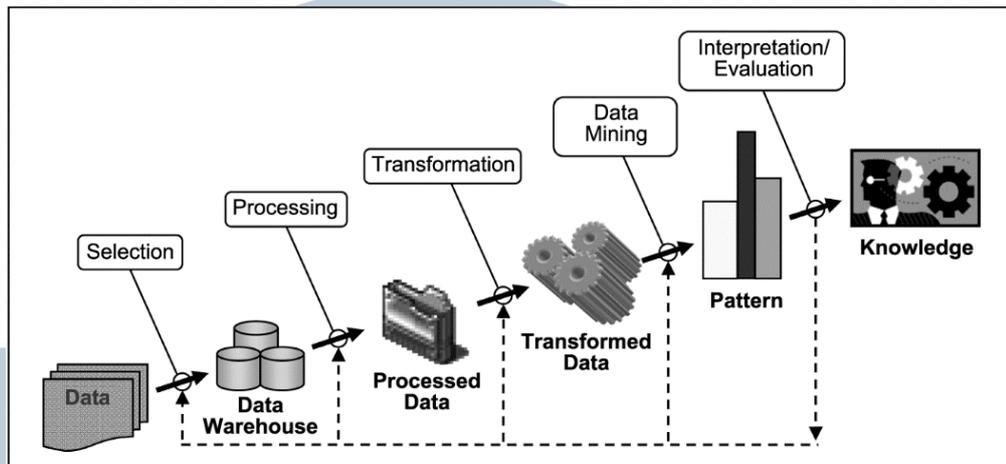
3.3. Metode Perancangan Visualisasi Data

Tabel 3.2. adalah perbandingan metode yang akan dipakai ketika ingin membuat visualisasi data.

Tabel 3.2. Perbandingan Metode Perancangan Visualisasi Data

<i>Knowledge Discovery in Databases (KDD)</i> (Santos & Azevedo, 2005)	<i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i> (Santos & Azevedo, 2005)
Proses KDD adalah proses menggunakan metode DM untuk mengekstrak apa yang dianggap pengetahuan sesuai dengan spesifikasi tindakan dan ambang batas, menggunakan basis data bersama dengan <i>pre-processing</i> , <i>sub sampling</i> , dan <i>transformation database</i> yang diperlukan.	Urutan enam tahapan CRISP-DM tidak kaku, CRISP-DM sangat lengkap dan didokumentasikan. Semua tahapannya diatur, terstruktur, dan didefinisikan sebagaimana mestinya, sehingga memungkinkan suatu proyek bisa mudah dipahami atau direvisi.
Lebih cocok untuk digunakan pada penelitian yang bersifat umum.	Lebih cocok untuk digunakan pada keperluan bisnis.

Hasil dari proses *prediction* tersebut akan ditampilkan dalam bentuk visualisasi data pada sebuah *dashboard* untuk mengetahui jumlah tamu yang akan datang agar dapat mencapai tingkat efisiensi yang tinggi di bagian akomodasi dan pekerja. Menurut (Shafique & Qaiser, 2014), *Knowledge Discovery in Databases* (KDD) dapat digunakan untuk penelitian dan proyek yang sifatnya lebih umum dan ruang lingkungannya lebih lengkap dan akurat.



Gambar 3.2. Knowledge Discovery in Databases (KDD)

Metode KDD digunakan dalam penelitian ini karena tahapan-tahapannya sangat mudah untuk dimengerti bagi banyak orang. Tahapan-tahapan yang ada pada KDD, yakni:

1. *Data Selection*

Proses pengumpulan dan pemilihan data. Data-data yang dibutuhkan antara lain:

- a. Jumlah akomodasi yang dimiliki tiap provinsi.
- b. Rata-rata pekerja yang ada di setiap provinsi.
- c. Tamu per hari yang tercatat masuk dalam hotel pada tiap provinsi.

Pengumpulan data akan didapatkan dengan cara mengunduh *file* melalui fitur yang telah disediakan oleh situs Badan Pusat Statistik. *File* yang diberikan oleh BPS berupa data dalam format *excel*.

2. *Pre-processing/Cleaning*

Data yang sudah dikumpulkan akan dirapikan dan “dibersihkan”. Langkah-langkah yang dilakukan pada tahap ini yaitu:

- a. Menghapus baris dan kolom yang tidak digunakan.
- b. Menjabarkan nama kolom.

3. *Transformation*

Data yang sudah dirapikan 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 diolah dalam SPSS agar dapat menjadi rumus menggunakan metode *prediction* dengan algoritma *Multiple Linear Regression*.

5. *Interpretation/ Evaluation*

Data yang telah di-*mining*, pola-polanya dapat diterjemahkan pada tahap ini.

Informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dengan syarat mudah untuk dimengerti oleh pihak yang memerlukannya.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.4. Teknik Pengumpulan Data

Penelitian ini akan menggunakan data dari Badan Pusat Statistik. Data-data yang diperlukan antara lain:

- a. Jumlah akomodasi yang dimiliki tiap provinsi.
- b. Rata-rata pekerja yang ada di setiap provinsi
- c. Tamu per hari yang tercatat masuk dalam hotel pada tiap provinsi.

Data yang digunakan merupakan data pariwisata yang didapat dari Badan Pusat Statistik. Pengumpulan data akan dilakukan dengan cara mencari setiap data yang dibutuhkan yang ada pada Badan Pusat Statistik.

3.5. Teknik Pengolahan Data

3.5.1. *Tools* untuk *Pre-Processing/Cleaning* dan *Interpretation*.

Terdapat 2 *tools* yang paling sering dipakai untuk melakukan proses ini, dapat dilihat perbandingannya pada tabel 3.3.

Tabel 3.3. Perbandingan *Tools Pre-processing* dan *Interpretation*

Tableau (Bunte, 2018)	PowerBI (Bunte, 2018)
Menawarkan alat BI yang kuat untuk meningkatkan visualisasi dan penemuan data untuk semua jenis organisasi dan pengguna bisnis.	<i>Platform</i> analitik dan intelijen bisnis berbasis <i>cloud</i> yang menawarkan tinjauan lengkap dari data paling kritis organisasi.
Memungkinkan pengguna membuat 24 jenis visualisasi dasar	Pengguna dapat memilih berbagai macam jenis visualisasi sebagai <i>blueprints</i>
Menawarkan dukungan untuk ratusan konektor data termasuk <i>Online Analytical Processing</i> (OLAP) dan opsi <i>Big Data</i> serta opsi <i>cloud</i> .	Sangat mampu terhubung ke sumber eksternal pengguna termasuk SAP HANA, JSON, MySQL, dan banyak lagi.
<i>Subscription</i>	<i>Free</i>

Penelitian ini memakai Microsoft PowerBI karena memiliki jenis visualisasi yang beragam dan mudah untuk digunakan, maka dipilihlah Microsoft PowerBI untuk digunakan dalam proses *pre-processing data* dan *interpretation*.

3.5.2. *Tools* untuk Data Mining.

Terdapat 2 *tools* yang paling sering dipakai untuk melakukan proses ini, dapat dilihat perbandingannya pada tabel 3.4.

Tabel 3.4. Perbandingan *Tools* untuk Data Mining

IBM SPSS Statistics (Kate, 2018)	RapidMiner (Kate, 2018)
<i>Subscription</i>	<i>Free</i>
Mudah untuk digunakan	Sangat mudah untuk digunakan
Sedikit sulit untuk <i>setup</i>	Sangat mudah untuk <i>setup</i>
Bagus untuk digunakan untuk data <i>mining</i>	Sangat bagus digunakan untuk data <i>mining</i>
Hasil analisis sangat baik	Hasil analisis baik
Visualisasi kurang bagus	Visualisasi bagus
<i>Report Generation</i> sangat baik	<i>Report Generation</i> baik

Penelitian ini memakai IBM SPSS Statistics karena memiliki hasil analisis yang sangat baik dan dapat pembuatan *report* dengan sangat baik, maka dipilihlah IBM SPSS Statistics untuk digunakan dalam proses data *mining*.

