

BAB 2

TINJAUAN PUSTAKA

2.1 Landasan Teori

Tahap ini bertujuan untuk memberikan informasi dan teori – teori yang berkaitan dengan topik yang diangkat pada tugas akhir ini.

2.2 Visualisasi Data

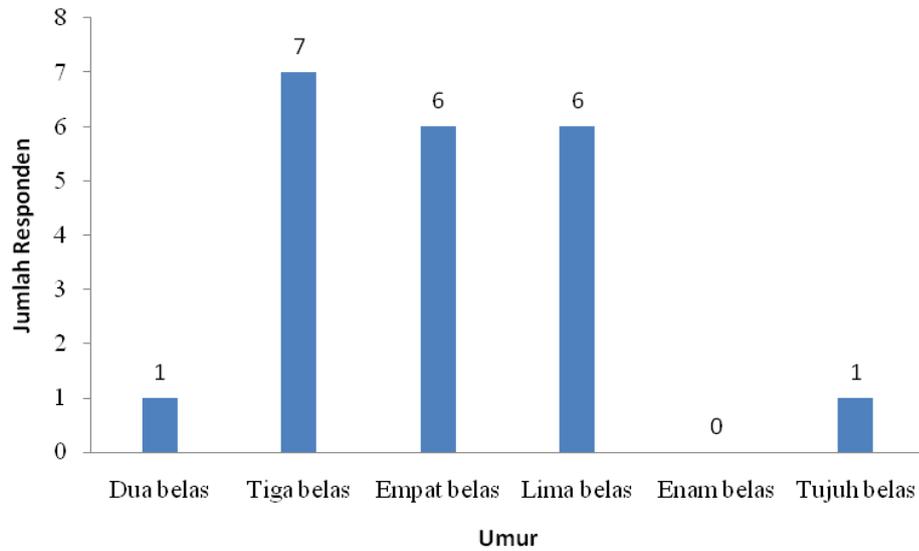
Visualisasi Data merupakan cara yang digunakan untuk memberikan gambaran terhadap suatu data yang ada dalam bentuk visual, seperti titik, garis dan tabung, batang. Sehingga lebih mudah dipahami, bentuk visual yang akan digunakan tidak boleh sembarangan, dan harus menggunakan diagram yang selaras dengan tipe data itu sendiri [4].

2.3 Tipe Tipe Visualisasi Data

Adapun contoh visualisasi data yang sering digunakan dalam menampilkan data adalah :

1. Diagram Batang

Diagram Batang merupakan tipe visualisasi data yang digunakan untuk perbandingan data. Diagram Batang memiliki beberapa tipe, diantaranya diagram horizontal, diagram vertical. Contoh diagram batang dapat dilihat dapat Gambar 2.1.



Gambar 2. 1 Diagram Batang

2. Tabel

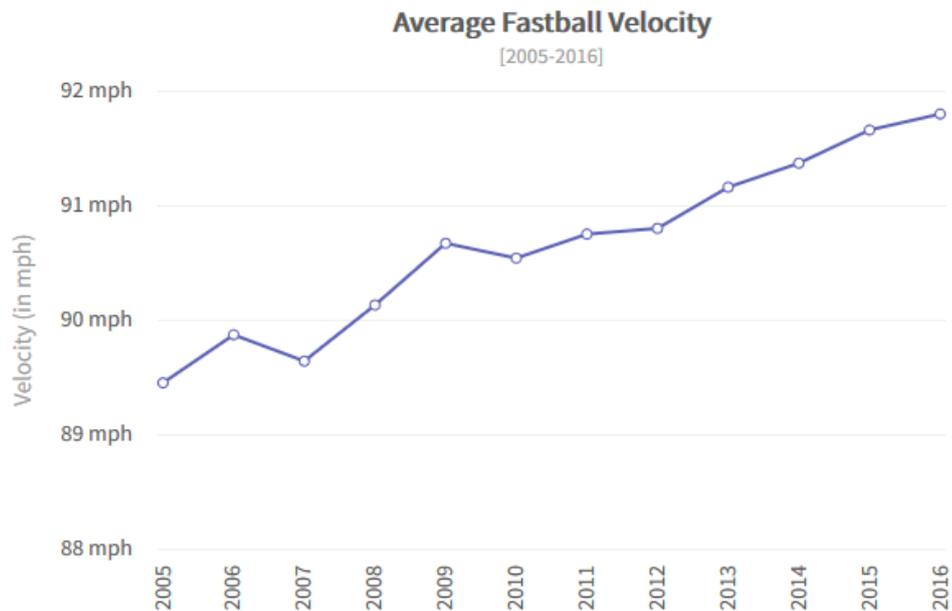
Tabel merupakan tipe visualisasi data yang digunakan untuk disampaikan kepada orang yang beragam dan memiliki fokus berbeda. Contoh tabel dapat dilihat pada Gambar 2.2.

	Name	Salary	Full Time Employee
1	Mike	\$10,000	✓
2	Jim	\$8,000	x
3	Alice	\$12,500	✓
4	Bob	\$7,000	✓

Gambar 2. 2 Tabel

3. Diagram Garis

Diagram Garis merupakan bentuk diagram yang digunakan untuk menggambarkan data yang berkesinambungan. Contoh diagram garis dapat dilihat pada Gambar 2.3.



Gambar 2. 3 Diagram Garis

2.4 Data

Data adalah suatu fakta mentah yang belum di olah dari sebuah peristiwa atau situasi yang ada, data yang diolah bisa menjadi tahapan awal untuk menentukan sebuah keputusan atau dasar pertimbangan dalam mengambil suatu kebijakan.yang ada[9].

2.5 Proses Visualisasi Data

Proses visualisasi data pada penelitian ini mengikuti tahapan yang dimuat dalam buku yang berjudul “Visualizing Data” [2]. Adapun tahapan dalam visualisasi data menurut buku tersebut sebagai berikut :

1. *Aquire*

Tahap ini adalah tahap dimana data dikumpulkan dari berbagai sumber. Tahap ini menjelaskan bagaimana suatu data diperoleh, apakah data diambil dari sebuah berkas atau data berasal dari internet.

2. *Parse*

Tahap ini adalah tahap dimana data dikemas dalam struktur tertentu dan dikelompokkan kedalam suatu kategori.

3. *Filter*

Tahap ini adalah proses dimana data yang tidak berhubungan dengan informasi yang dibutuhkan akan disampaikan untuk dihilangkan.

4. *Mine*

Tahap ini adalah tahap dimana menggunakan metode statistika atau data mining untuk jalan mencari pola atau dijabarkan kedalam konteks matematis.

5. *Represent*

Tahap ini adalah tahap dimana data diubah menjadi data yang berbentuk visual dasar, seperti tabel, bar, *list* atau *tree*.

6. *Refine*

Pada tahap ini adalah tahap dimana model visual dibuat lebih jelas lagi dan menarik.

7. *Interact*

Tahap ini adalah dimana penggunaan metode untuk memanipulasi data ditambahkan, atau pengguna dapat menampilkan data yang diinginkan.

2.6 Basis Data

Basis Data adalah sekumpulan data yang saling berhubungan satu sama lain, yang diorganisir sedemikian rupa yang nantinya akan digunakan kembali dengan cepat. Tujuan dalam adanya basis data adalah untuk mempercepat pengambilan data, efisiensi ruang penyimpanan, akurat, ketersediaan, kelengkapan, keamanan dan bersifat multiuser. Adapun sistem yang akan mengelola data tersebut biasa disebut DBMS (*Database Management System*)[9].

2.7 Statistika Deskriptif

Statistika deskriptif adalah salah satu ilmu yang menjelaskan tahapan dalam pengumpulan, mengolah, penyajian dan menganalisis data kuantitatif secara

deskriptif, sehingga menghasilkan informasi yang lebih jelas[20]. Adapun contoh penyajian data bisa berupa : Tabel, *Histogram*, *Pie Chart*.

2.8 Search Engine Optimization

Search Engine Optimization atau sering disingkat *SEO* merupakan teknik pemasaran online yang bertujuan mempromosikan unit bisnis yang ada di internet, *Search Engine Optimization* merupakan salah satu dari dua tipe metode dalam *Search Engine Marketing*, yang sering dan banyak dilakukan adalah *Search Engine Optimization* dan *Paid-Per-Click* [8]. *Search Engine Optimization* merupakan cara untuk mendapatkan audience kehalaman web dari *search engine* berdasarkan *keyword* yang di cari oleh *audience*[6], dimana dalam prosesnya, para pemilik bisnis atau *webmaster* berlomba-lomba untuk menjadi nomor satu di halaman pertama google berdasarkan *keyword* yang sering dicari atau *keyword* yang menjadi fokus pada bisnis mereka[13]

2.9 Data Mining

Data Mining adalah tahapan menemukan pola dari data mentah dalam artian yang belum diketahui makna atau kegunaannya, yang tersimpan dalam jumlah yang besar, nantinya akan di implementasikan atau di proses menggunakan ilmu statistik

dan matematika untuk mencari informasi yang berguna[12] Pada dasarnya, data mining berhubungan dengan analisa data dan mencari pola dari jumlah data yang besar atau dari sumber yang tidak diduga. Adapun langkah-langkah yang ada dalam proses data mining sebagai berikut :

1. *Data Cleaning*

Untuk menghilangkan noise data yang tidak konsisten.

2. *Integration*

Menggabungkan beberapa file ada basis data.

3. *Selection*

Data yang berkaitan dengan proses analisis dikembalikan kedalam basis data untuk dilakukannya proses data mining.

4. *Transformation*

Data yang berubah bentuk atau menjadi kesatuan yang menghasilkan bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresif.

5. *Mining*

Adalah metode yang digunakan untuk mengekstrak pola dari sebuah data.

6. *Knowledge discovery*

Yaitu proses ditemukannya sebuah pola yang terdapat pada sebuah data.

7. *Pattern evolution*

Mengidentifikasi pola yang menarik yang mewakili pengetahuan berdasarkan beberapa tindakan yang menarik.

8. *Knowledge presentation*

Menambahkan teknik visualisasi dalam pengetahuan untuk memberikan informasi lebih kepada user.

2.10 Restfull Web Services

Restfull Web Services adalah standarisasi sebuah arsitektur dalam pengembangan suatu web aplikasi. Rest biasa digunakan untuk berkomunikasi antar data menggunakan protocol HTTP / HTTPS. Untuk bisa berkomunikasi dengan data, maka Restfull bekerja bersamaan dengan API dan keduanya disebut dengan *Web Services*. Rest terbagi menjadi dua komponen, ada namanya *Rest Client* dan *Rest Server*, dimana Rest Server digunakan untuk melayani request pada data melalui http request sedangkan Rest Client digunakan untuk meminta resources kepada server.

2.11 Analisis Dan Design Berorientasi Objek

Analisis dan Desain Berorientasi Objek (*Object Oriented Analysis and Design*) adalah cara baru dalam memecahkan suatu masalah dengan model yang dibuat menurut konsep.

2.12 JSON

JSON (*Javascript Object Notation*) adalah standar untuk mempresentasikan data sebagai atribut dengan nilai. JSON berasal dari syntax Javascript untuk digunakan dalam aplikasi web sebagai alternatif yang lebih ringan dan terstruktur dibandingkan dengan XML (*Extensible Markup Language*). JSON memberikan cara yang baik untuk merangkum data antara klien dan server. Adapun kriteria JSON sebagai berikut:

1. Penyusunan data dalam atribut dan nilai, masing masing dipisahkan oleh titik dua (:).
2. Atribut muncul sebagai karakter string dan diapit dengan tanda petik (“) sebelum titik dua (:).
3. Nilai yang ada sebelum titik dua (:) bisa berupa karakter string, angka, array, object dan dapat diisi juga dengan nilai boolean.

2.13 Usability Testing

Usability testing adalah metode evaluatif yang memungkinkan tim untuk mengamati pengalaman individu dengan aplikasi digital. Metode ini dirancang untuk membantu identifikasi bagian antarmuka yang paling membingungkan orang sehingga dapat diperbaiki dan diuji ulang sebelum diluncurkan[18].

2.14 Algoritma DBSCAN

Algoritma *Density-based spatial clustering of applications with noise* (DBSCAN) adalah algoritma *clustering* yang sering digunakan dalam data mining dan *machine learning*. Metode ini menganggap cluster sebagai suatu area yang berisi objek-objek yang padat atau sesak, yang dipisahkan oleh area yang memiliki kepadatan rendah (merekpresentasikan *noise*)[7].

Algoritma ini, menumbuhkan area-area dengan kepadatan yang cukup tinggi ke dalam cluster-cluster and menemukan cluster-cluster dalam bentuk yang sembarang dalam suatu database spatial yang memuat noise (Sander et al., 1998). DBSCAN mendefinisikan cluster sebagai himpunan maksimum dari titik-titik kepadatan yang terkoneksi (*density-connected*). Semua objek yang tidak masuk ke

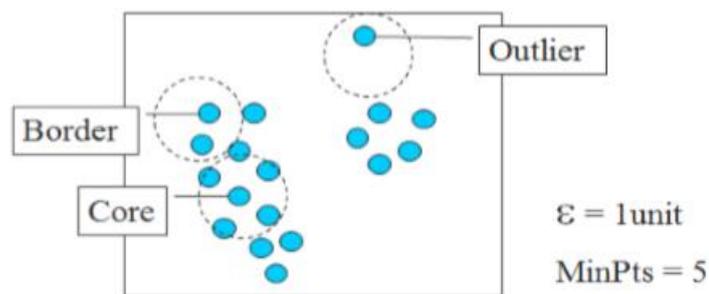
dalam cluster manapun dianggap sebagai noise.

2.14.1 Konsep Kepadatan (*Density Concept*)

DBSCAN menentukan sendiri jumlah cluster yang akan dihasilkan sehingga kita tidak perlu lagi untuk menentukan jumlah cluster yang diinginkan, tapi memerlukan 2 input lain[19], yaitu:

- a. *MinPts*: minimal banyak items dalam suatu cluster
- b. *Eps*: nilai untuk jarak antar-items yang menjadi dasar pembentukan neighborhood dari suatu titik item.

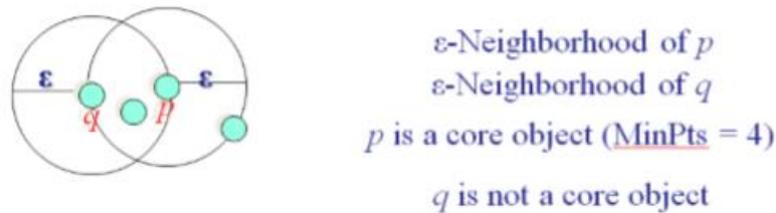
Neighborhood yang terletak di dalam radius(ϵ) disebut ϵ -neighborhood dari objek data. Jika ϵ -neighborhood dari suatu objek berisi paling sedikit suatu angka yang minimum, *MinPts* dari suatu objek, objek tersebut disebut *core object*. Neighborhood dari border points berisi jauh lebih sedikit items daripada neighborhood dari core points. Suatu border point bisa jadi termasuk ke dalam lebih dari 1 core object. Berikut ini gambar yang menunjukkan mana yang merupakan border point dan mana yang merupakan core point contoh dengan menggunakan *MinPts* = 5 dan *Eps*=1. Bentuk dari Core dan Border dapat dilihat pada Gambar 2.4.



Gambar 2. 4 Core and Border

Menurut definisi, ada 2 jenis titik (points) dalam suatu cluster: di dalam cluster (core points) dan di tepian cluster (border points) di mana neighborhood dari border points berisi jauh lebih sedikit items daripada neighborhood dari core

points (Ester et al., 1996). Suatu border point bisa jadi termasuk ke dalam lebih dari 1 cluster.



Gambar 2. 5 Konsep Kepadatan

DBSCAN menelusuri cluster-cluster dengan memeriksa ϵ neighbourhood (Eps-neighborhood) dari tiap-tiap point dalam database. Jika ϵ neighborhood dari point p mengandung lebih dari $MinPts$, cluster baru dengan p sebagai core object diciptakan.

Kemudian DBSCAN secara iteratif mengumpulkan secara langsung objek-objek density-reachable dari core object tersebut, dimana mungkin melibatkan penggabungan dari beberapa cluster-cluster yang density-reachable.

2.14.2 *Directly Density-Reachable*

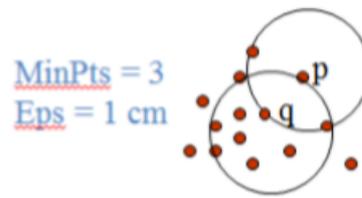
Sebuah titik item dikatakan *directly density-reachable* dari titik lainnya jika jarak di antara mereka tidak lebih dari nilai Eps. Directly density-reachable = titik q dikatakan directly density-reachable dari titik p jika titik q adalah $\in N_{eps}(p)$ dan p adalah core point.

$$N_{Eps}(p) = \{ q \text{ belong to } D \mid dist(p, q) \leq Eps \}$$

Jarak dari titik ke titik lainnya tidak lebih dari nilai Eps.

$$\text{Core Object } p : |N_{eps}(p)| \geq MinPts$$

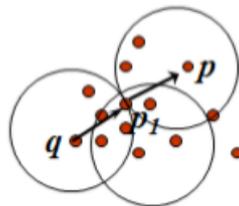
Adapun bentuk Directly Density-Reachable dapat dilihat pada Gambar 2.6



Gambar 2. 6 Directly Density-Reachable

2.14.3 *Density-Reachable*

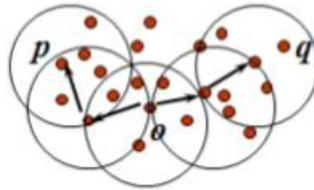
Sebuah titik item dikatakan density-reachable dari titik item yang lain jika ada suatu rantai yang menghubungkan keduanya yang berisi hanya titik-titik yang directly density-reachable dari titik-titik sebelumnya. Suatu objek p adalah density reachable dari objek q dengan respek ke ϵ dan MinPts dalam suatu set objek D jika terdapat suatu rantai objek p_1, p_2, \dots, p_n , dimana $p_1 = q$ dan $p_n = p$, di mana p_{i+1} density reachable secara langsung dari p_i dengan respek ke ϵ dan MinPts, untuk $1 \leq i \leq n$, p_i anggota D . Adapun bentuk Density-Reachable dapat dilihat pada Gambar 2.7.



Gambar 2. 7 Density-Reachable

2.14.4 *Density-Connected*

Sebuah obyek p adalah density-connected terhadap obyek q dengan memperhatikan Eps dan MinPts dalam set obyek D , jika ada sebuah obyek o elemen D sehingga p dan q keduanya density-reachable dari o dengan memperhatikan Eps dan MinPts.



Gambar 2. 8 Density-Connected

2.14.5 Cluster

Misalkan D adalah sebuah database dari poin. Sebuah Cluster C wrt. ϵ dan MinPts adalah himpunan tidak kosong dari D memenuhi kondisi berikut:

- 1) $\forall p, q$: jika $p \in C$ dan q adalah density-reachable dari p wrt. ϵ dan MinPts , maka $q \in C$. (Maximality)
- 2) $\forall p, q \in C$: p adalah density-connected ke q wrt. ϵ dan MinPts . (Connectivity).

2.14.6 Noise

Misalkan C_1, \dots, C_k . Menjadi cluster dari Database D wrt. parameter ϵ dan MinPts , $i = 1, \dots, k$. Kemudian kita mendefinisikan noise sebagai himpunan titik-titik dalam database D tidak dimiliki pada C_i cluster, yaitu $\text{noise} = \{p \in D \mid \forall i: p \notin C_i\}$.