

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Data Governance**

*Data Governance* atau tata kelola data adalah organisasi dan implementasi kebijakan, prosedur, struktur, peran, dan tanggung jawab yang menguraikan dan menegakkan aturan keterlibatan, hak keputusan, dan akuntabilitas untuk pengelolaan aset informasi yang efektif [7]. Tata kelola data mengacu pada pelaksanaan otoritas dan kontrol atas pengelolaan data.

#### **2.2 Pengangguran**

Pengangguran adalah suatu keadaan di mana seseorang yang tergolong dalam angkatan kerja ingin mendapatkan pekerjaan tetapi belum dapat memperolehnya [8]. Pengangguran meliputi penduduk yang tidak bekerja tetapi sedang mencari pekerjaan, atau mempersiapkan suatu usaha, atau merasa tidak mungkin mendapat pekerjaan (putus asa), atau sudah diterima bekerja tetapi belum mulai bekerja [9].

Data mengenai pengangguran dapat diolah menjadi sebuah informasi yang mampu dimengerti oleh masyarakat umum maupun kepada pemerintah. Agar informasi yang dihasilkan dapat lebih dipahami dengan baik, data tersebut diolah menjadi informasi berbentuk visual.

#### **2.3 COVID-19**

*Novel Coronavirus* atau yang diberi nama 2019-nCoV dan dikenal dengan nama COVID-19 pertama kali muncul di Wuhan, Cina pada akhir tahun 2019.

COVID-19 telah menyebar ke berbagai negara didunia dan telah menginfeksi banyak orang dan semakin hari semakin banyak kasus yang dihasilkan dari penularan antar manusia. Virus ini menyebar melalui batuk dan bersin tanpa menutupi mulut yang dapat menyebarkan tetesan ke udara, menyentuh atau berkontak langsung dengan orang yang terkena virus dan melakukan kontak dengan permukaan atau benda yang tertular virus kemudian menyentuh hidung, mata atau mulut. National Institutes of Health (NIH) menyatakan bahwa ada beberapa kelompok orang yang memiliki risiko tertinggi terinfeksi COVID-19 yaitu anak-anak, lansia berusia 65 atau lebih tua dan ibu hamil. Pencegahan yang dapat dilakukan adalah dengan berada di dalam rumah, tidak berkontak langsung dengan orang lain, menutup mulut dan hidung saat batuk dan bersin dan menjaga kebersihan. Karena kurangnya vaksin yang efektif, tindakan terbaik yang dilakukan untuk mengendalikan virus pada manusia adalah memperkuat pengendalian sistem pengawasan kesehatan masyarakat dan melakukan pengujian diagnostik serta melakukan karantina”[10].

Pandemi COVID-19 di berbagai negara seperti Rusia, Prancis, Itali, India, Jerman, Kanada, Malaysia, Peru, Iran, Turki dan negara lainnya. Indonesia menjadi salah satu negara yang terdampak akibat COVID-19. Indonesia berada di urutan ke-20 jumlah kasus terbanyak di dunia [11]. Berbagai sektor menjadi terdampak salah satunya adalah ketenagakerjaan dan mengakibatkan adanya pengangguran.

#### **2.4 Visualisasi Data**

Visualisasi merupakan mekanisme bagaimana manusia mempersepsi dan menggunakan informasi visual sehingga informasi melalui grafis tersebut

tersampaikan secara singkat dan jelas [12]. Pada umumnya, visualisasi digunakan untuk mengagresi data dalam jumlah yang sangat besar yang kemudian dipresentasikan dengan berbagai model [13]. Informasi yang ditampilkan secara visual akan lebih mudah dimengerti dan dipahami oleh pembaca. Kemampuan analisa manusia terhadap data yang dipresentasikan dalam bentuk visual lebih baik daripada teks. Hasil dari visualisasi tersebut ditampilkan dalam bentuk *Dashboard* yang dapat digunakan untuk menampilkan informasi, menganalisis data dan memantau [14].

## **2.5 Dashboard**

*Dashboard* adalah tampilan visual dari sebuah informasi penting yang diperlukan untuk mencapai satu atau lebih tujuan dan diatur dan terkonsolidasi dalam sebuah layer sehingga informasi dapat dipantau dalam sekejap [15]. Dalam *dashboard*, informasi disajikan dalam bentuk visual dengan kombinasi tulisan dan gambar yang merepresentasikan sebuah data.

Terdapat 3 tipe *dashboard*, yaitu [16]:

1. *Strategic Dashboard*, digunakan untuk mendukung manajemen tingkat strateegis dalam memberikan informasi yang dibutuhkan untuk membuat keputusan bisnis, memprediksi peluang dan memberikan arahan untuk mencapai tujuan strategis.
2. *Tactical Dashboard*, digunakan untuk mendukung kemajuan proyek dan analisis untuk menentukan penyebab kondisi atau peristiwa tertentu.

3. *Operational Dashboard*, digunakan untuk mendukung pemantauan proses bisnis dan kegiatan bisnis yang kompleks. *Dashboard* ini memberikan pembaruan harian atau mingguan berbentuk grafik *real time*.

Untuk dapat menghasilkan sebuah *Dashboard* tentunya akan dilakukan proses pembuatan visualisasi ditahap sebelumnya. Terdapat berbagai jenis *tools* yang dapat digunakan dalam pembuatan visualisasi yaitu Power BI, QlikView, Qlik Sense, Tableau dan lain-lain. Tableau merupakan salah satu *tools* yang digunakan dalam pembuatan visualisasi serta menyediakan berbagai fitur yang diperlukan untuk pemrosesan data hingga mempresentasikan laporan dan fitur *drag and drop* yang dapat memudahkan pengguna.

## **2.6 Tableau**

Tableau adalah sebuah platform analitik visual yang mengubah cara dalam menggunakan data untuk memecahkan sebuah masalah seseorang atau organisasi untuk memaksimalkan data yang mereka miliki [17]. Tableau menyediakan fasilitas untuk dapat terhubung ke berbagai sumber data dengan berbagai tipe seperti sistem data yang diatur dalam format file CSV, JSON, XML, Ms. Excel, dll, sistem data relasional dan nonrelasional seperti PostgreSQL, MySQL, SQL Server, MongoDB, dll, sistem cloud seperti AWS, Oracle Cloud, Google BigQuery dan Microsoft Azure[18].

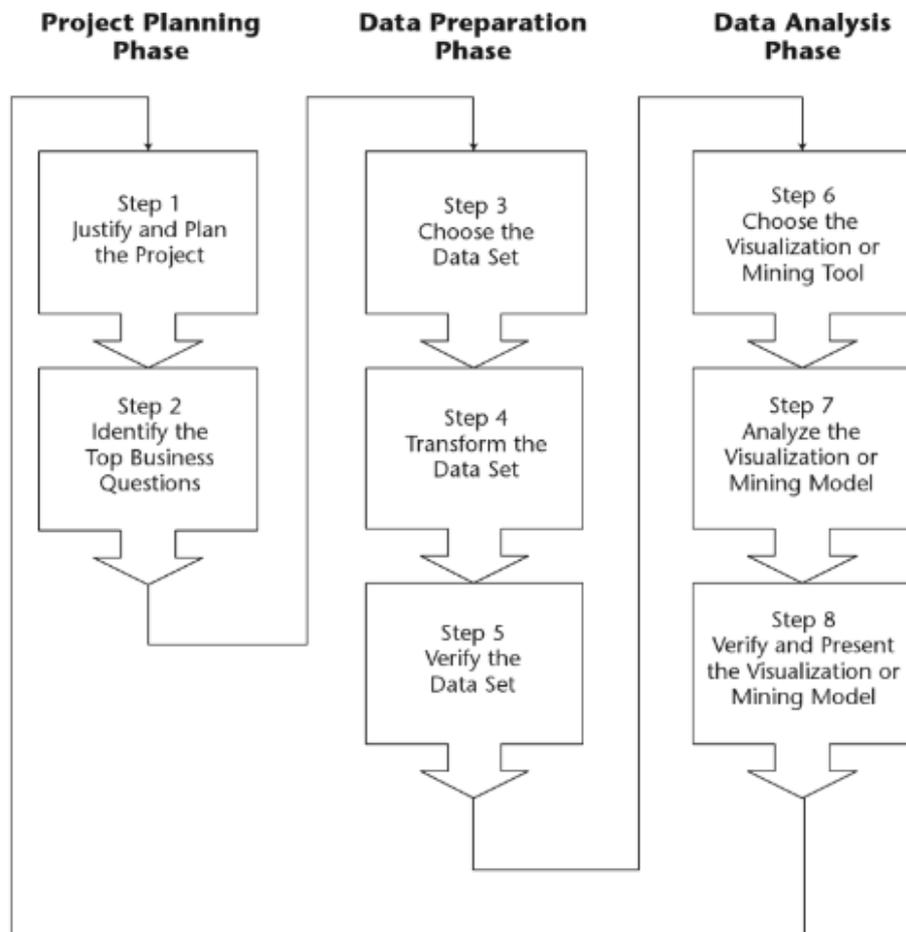
Tableau menghasilkan informasi visualisasi berdasarkan data yang dimiliki. Sebelum visualisasi dihasilkan, tentunya dilakukan berbagai tahapan analisa untuk dapat menghasilkan informasi yang tepat dan tidak menyimpang. Terdapat berbagai

metode yang digunakan dalam proses analisa untuk dapat menghasilkan sebuah informasi visualisasi, salah satunya adalah *Visual Data Mining*.

## **2.7 Visual Data Mining (VDM)**

*Visual data mining* (VDM) adalah proses interaksi dan penalaran analitis dengan satu atau lebih representasi visual dari data abstrak. Proses tersebut dapat mengarah pada penemuan visual dari pola yang kuat dalam data ini atau memberikan beberapa panduan untuk penerapan teknik data mining dan analitik lainnya [19]. *Visual data mining tools and techniques* membantu dalam membuat visualisasi dari model *data mining* untuk mendapatkan pengetahuan dan wawasan terhadap pola berdasarkan algoritma *data mining* yang dapat membantu dalam membuat keputusan dan memprediksi kesempatan baru dalam bisnis [20]. Ada 3 fase dalam data visualization dan visual data mining (VDM) [20]:

1. *Project Planning*
2. *Data Preparation*
3. *Data Analysis*



**Gambar 2. 1 Eight-Step Data Visualization and Visual Data Mining Methodology**

### **2.7.1 Project Planning Phase**

#### **1. Justify and plan the project**

Pada tahap ini, merupakan tahap pengenalan metode pertama dari delapan tahap dalam visualisasi data dan visual data mining (VDM). Dalam beberapa kasus, proyek itu sendiri membutuhkan beberapa justifikasi bisnis sebelum dimulai (mendapatkan pendanaan untuk proyek). Tahap ini juga mendiskusikan perencanaan proyek VDM dan menyediakan panduan dalam mengestimasi waktu proyek dan kriteria sumber yang dibutuhkan dalam proyek tersebut dan akan membantu

dalam menemukan peran tim dan tanggung jawab yang akan dilakukan dalam proyek.

## *2. Identifying the Top Business Questions*

Pada tahap ini, merupakan pengenalan tahap kedua dari metode VDM yaitu bagaimana cara menentukan pertanyaan dalam bisnis sehingga mampu untuk dilakukan investigasi atau penyelidikan terhadap visualisasi data dan VDM. Dalam tahap ini juga dapat membantu dalam memetakan pertanyaan bisnis ke dalam visualisasi data dan mendefinisikannya ke dalam *visual data mining problem*.

### **2.7.2 Data Preparation Phase**

#### *1. Choosing the Data*

Pada tahap ini, merupakan pengenalan tahap ketiga dari metode VDM dan mendiskusikan pemilihan data terkait dengan pertanyaan dalam visualisasi data dan VDM di tahap sebelumnya. Dalam tahap ini juga memperkenalkan konsep menggunakan eksplorasi data mart sebagai tempat penyimpanan (repository) untuk membangun dan mempertahankan data sets bisnis yang diselidiki.

#### *2. Transforming the Data Set*

Pada tahap ini, merupakan pengenalan tahap keempat dari metode VDM dan mendiskusikan cara menampilkan logical transformation dalam dataset bisnis yang tersimpan dalam exploratory data mart. Tahap ini membantu dalam mengubah dataset untuk proyek VDM dengan

mempresentasikan dan mendiskusikan berbagai contoh praktis dan diaplikasikan ke tahap berikutnya.

#### 4. *Verifying the Data Set*

Pada tahap ini, merupakan pengenalan tahap kelima dari metode VDM dan mendiskusikan bagaimana cara untuk melakukan verifikasi produksi dataset bisnis yang berisikan expected data dan semua data dari tahap tiga dan tahap empat yang sudah benar, tidak ada error dalam dataset.

#### 5. *Choosing the Visualization or Data Mining Tool*

Pada tahap ini, merupakan pengenalan tahap keenam dari metode VDM dan mendiskusikan cara untuk memilih model tool visualisasi data atau data mining yang sesuai dalam menyelidiki pertanyaan bisnis.

### **2.7.3 Data Analysis Phase**

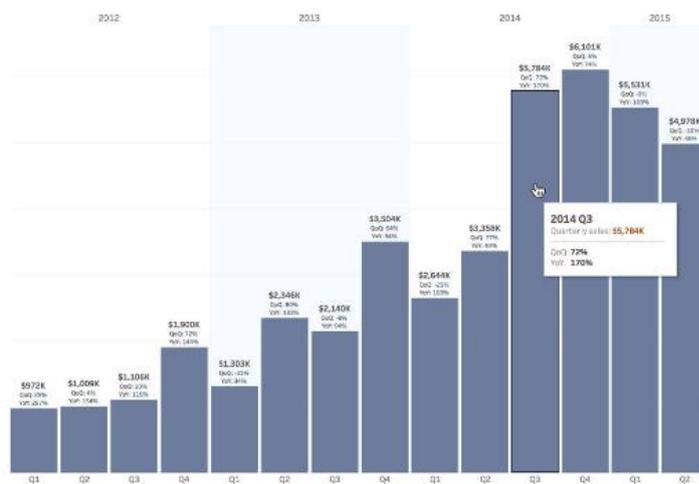
#### 1. *Analyzing the Visualization or Data Mining Model*

Pada tahap ini, merupakan pengenalan tahap ketujuh dari metode VDM dan mendiskusikan cara menggunakan model visualisasi data atau data mining untuk mendapatkan pengetahuan yang mendalam dalam menjawab pertanyaan bisnis yang telah diidentifikasi. Setiap model dapat diselidiki secara visual untuk menemukan pola atau pattern (tren bisnis).

Ada beberapa tipe Chart dan Graph yang digunakan dalam pembuatan visualisasi untuk melakukan analisa dan mengolah data set yang ada untuk menghasilkan sebuah informasi [21]:

a. *Bar Chart*

*Bar Chart* adalah satu dari banyak data visualization yang paling sering digunakan untuk dapat melakukan perbandingan antar kategori, perbedaan highlight, menampilkan tren dan outliers, dan menampilkan angka tertinggi dan terendah. *Bar Chart* akan lebih efektif ketika data yang dimiliki dapat dipisah menjadi beberapa kategori.



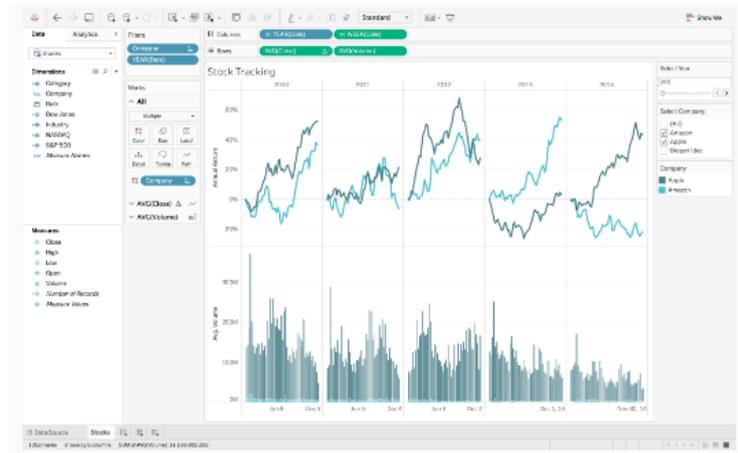
Gambar 2. 2 Bar Chart

Sumber: [21]

b. *Line Chart*

*Line Chart* menghubungkan beberapa titik data yang berbeda dan disajikan untuk perubahan yang berkelanjutan. Bentuk

diagram ini biasanya digunakan untuk melihat tren data yang terkait dengan waktu.

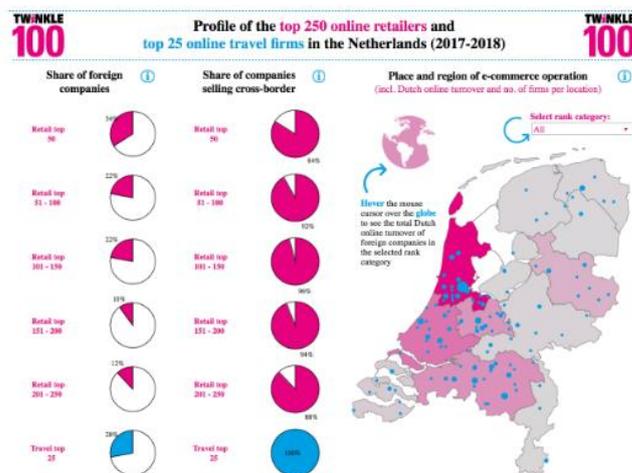


**Gambar 2. 3 Line Chart**

Sumber: [21]

c. Pie Chart

Pie Chart berguna untuk menambahkan detail pada visualisasi lain. Diagram ini dapat digunakan untuk membuat perbandingan antar kolom dalam satu visual.

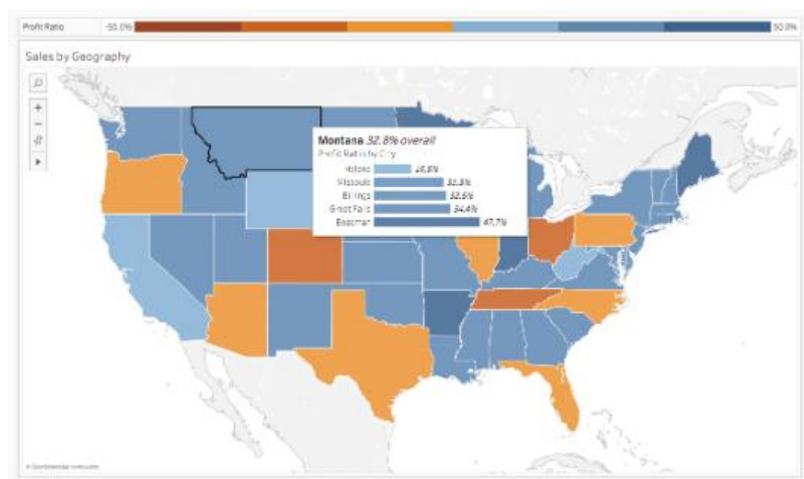


**Gambar 2. 4 Pie Chart**

Sumber: [21]

d. *Maps*

*Maps* digunakan untuk membuat visualisasi berdasarkan informasi lokasi yang terdapat dalam data seperti kode pos, nama kota, nama negara, dll. Jika dalam sebuah data terdapat informasi geografis, maps dapat dengan mudah menampilkan lokasi yang sesuai berdasarkan tren yang dimiliki oleh data.

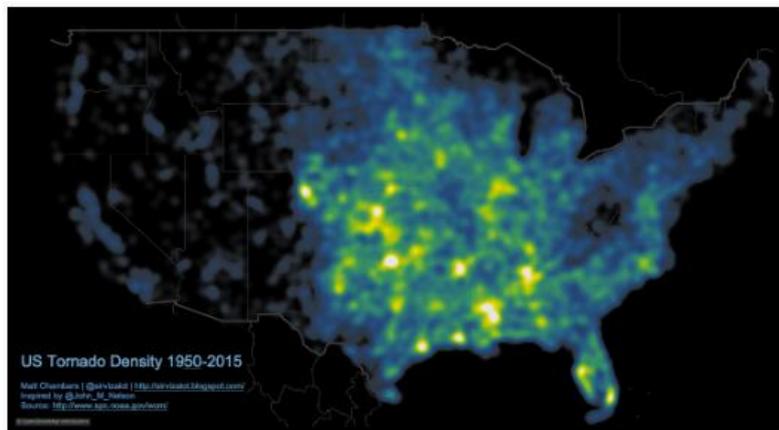


**Gambar 2. 5 Maps**

**Sumber: [21]**

e. *Density Maps*

*Density Maps* menampilkan pola dan mengidentifikasi lokasi dengan jumlah titik data yang lebih atau lebih sedikit. *Density Maps* akan lebih efektif terhadap kumpulan data yang berisi banyak titik data di area geografis.

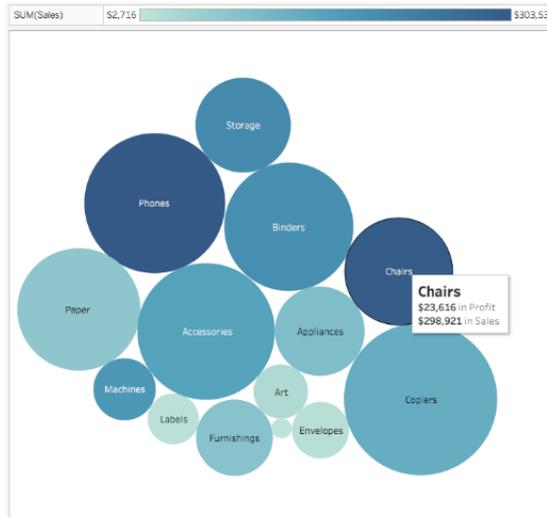


**Gambar 2. 6 Density Maps**

**Sumber: [21]**

f. *Bubble Chart*

*Bubble Chart* menampilkan hubungan antar value yang ada di dalam data berdasarkan ukuran dari lingkaran dan dapat mempresentasikan jumlah data yang besar dalam satu visual

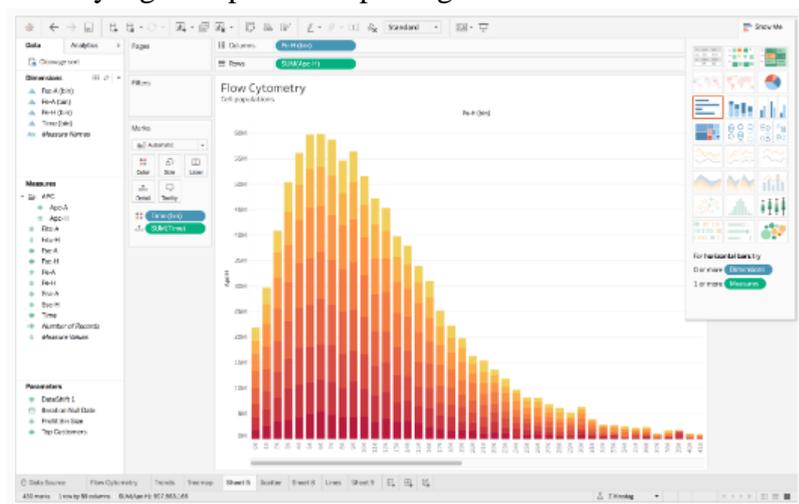


**Gambar 2. 7 Bubble Chart**

**Sumber: [21]**

g. Histogram Chart

Histogram menunjukkan bagaimana data didistribusikan ke dalam sebuah kelompok dan menggolongkan data yang dimiliki ke dalam kategori yang lebih spesifik dan ditetapkan ke dalam sebuah bentuk bar yang disesuaikan dengan jumlah data yang terdapat disetiap kategori.

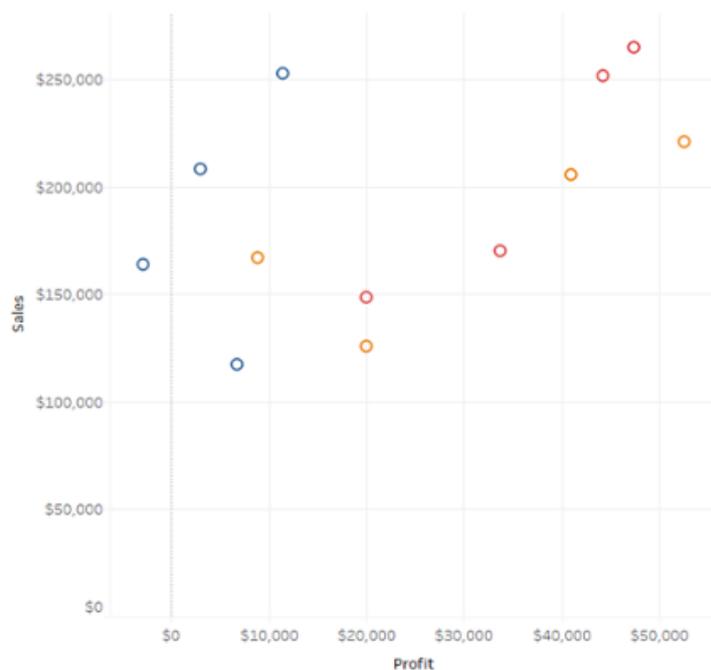


**Gambar 2. 8 Histogram Chart**

**Sumber: [21]**

#### h. Scatter Plot

Scatter Plot merupakan visualisasi yang digunakan untuk membandingkan dua *measure*. Berbagai aspek dari data yang ada dapat ditampilkan melalui banyak bentuk, warna dan ukuran. Selain itu dapat juga dilakukan penambahan berbagai aspek untuk dapat menampilkan korelasi dari kedua *measure*.



**Gambar 2. 9 Scatter Plot**

**Sumber: [21]**

#### 2. Verifying and Presenting Analysis

Pada tahap ini, merupakan pengenalan tahap akhir dari metode VDM dan mendiskusikan tentang melakukan verifikasi bahwa visualisasi dan model *data mining* yang telah dibuat sudah mencapai tujuan dan objek bisnis, mempresentasikan hasil penemuan visualisasi dan *data mining* kepada pembuat keputusan.

## 2.8 Korelasi Pearson

Korelasi Pearson adalah analisis dalam statistik untuk dapat melihat hubungan linear antara dua variabel yang memiliki skala data interval atau rasio [22].

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}$$

**Rumus 2.1 Korelasi Pearson**

**Sumber: [22]**

Rumus 2.1 merupakan rumus korelasi pearson, koefisien korelasi dapat memiliki nilai positif atau negatif dalam rentang nilai -1 dan 1. Apabila nilai mendekati -1 atau 1 maka hubungan antara dua variabel semakin kuat. Namun, apabila nilai mendekati 0 maka hubungan antara dua variabel semakin lemah. Interpretasi nilai koefisien korelasi adalah sebagai berikut:

1. 0 – 0.2 adalah hubungan antar variabel sangat lemah
2. 0.2 – 0.4 adalah hubungan antar variabel lemah
3. 0.4 – 0.7 adalah hubungan antar variabel cukup kuat
4. 0.7 – 0.9 adalah hubungan antar variabel kuat
5. 0.9 – 1 adalah hubungan antar variabel sangat kuat

Dalam Tableau, terdapat rumus yang digunakan untuk mendapatkan nilai korelasi antar variabel yaitu dengan membuat *calculated field* dengan fungsi WINDOW\_CORR.

```
WINDOW_CORR(SUM([Profit]), SUM([Sales]))
```

## Rumus 2.2 Rumus Korelasi Pearson di Tableau

Sumber: [23]

Rumus 2.2 merupakan rumus korelasi yang digunakan dalam Tableau dengan menggunakan fungsi WINDOW\_CORR yang dibuat melalui *calculated field*. Profit dan Sales dalam rumus disesuaikan dengan variabel yang ingin dihitung angka korelasinya. Dalam penelitian ini, variabel yang digunakan adalah kasus, kematian dan kesembuhan COVID-19 dan tingkat pengangguran.

### 2.9 Penelitian Terdahulu

Berikut ini merupakan beberapa referensi dari penelitian terdahulu yang terkait dengan metode serta penulisan dalam penelitian ini.

Tabel 2. 1 Penelitian Terdahulu

Judul Penelitian	Pembahasan
<p><i>Visualization and Prediction of Film Award Nominations by Using of Visual Data Mining (VDM) and Exploratory Data Analysis</i></p> <p>Peneliti:</p> <ol style="list-style-type: none"><li>1. Rayhanali Heiko Amier</li><li>2. Johan Setiawan</li></ol> <p>Nama Jurnal: IEEE Xplor</p> <p>2019 5<sup>th</sup> International Conference on New Media Studies (CONMEDIA)</p> <p>Electronic ISBN:978-1-7281-0726-4</p>	<p>Penelitian ini bertujuan untuk melakukan perancangan visualisasi tentang nominator <i>Best Picture</i> dalam Academy Awards periode 1993-2017 dalam bentuk <i>dashboard</i> dengan menggunakan metode <i>Visual Data Mining</i> (VDM) dan menerapkan metode prediksi Holt-Winters <i>exponential</i> smoothing untuk memprediksi dari beberapa nilai parameter yang diteliti untuk tahun 2018 serta menggunakan <i>tools</i> Tableau. Data diambil melalui internet dari tiga sumber.</p> <p>Hasil dari penelitian ini adalah membuat 5 <i>dashboard</i> yang menampilkan pencarian pola tren terhadap parameter yang diteliti dan</p>

<p>USB ISBN:978-1-7281-0725-7 Print on Demand (PoD) ISBN:978-1-7281-1574-0</p>	<p>berhasil untuk menerapkan metode prediksi. Penelitian ini juga melakukan valiasi dengan melakukan pengujian <i>User Acceptance Test (UAT)</i> terhadap 5 responden.</p>
<p>Visualisasi Data Analisa Keterangan Rencana Kota (KRR) di Kota Bandung Tahun 2013-2017</p> <p>Peneliti: Intan Saktia Rahmawati</p> <p>Sumber: kc.umh.ac.id</p>	<p>Tujuan penelitian ini adalah mengetahui perbandingan dari tingkat densitas jumlah permohonan masuk dari berbagai wilayah yang memiliki jumlah tertinggi dan terendah dan ditampilkan dalam bentuk visual.</p> <p>Metode yang digunakan adalah <i>Visual Data Mining (VDM)</i> dan <i>tools</i> Tableau Desktop.</p> <p>Hasil dari penelitian ini yaitu mengetahui grafik dan presentase dari tingkat densitas permohonan terkait izin mendirikan bangunan 5 tahun terakhir periode tahun 2013-2017 di Kota Bandung. Permohonan tertinggi yaitu wilayah Arcamanik sebesar 17,35% dengan total permohonan 4.357 dan permohonan tertinggi yaitu wilayah Gedebage sebesar 6,61% dengan total permohonan 1.660.</p>
<p>Visualisasi Data Terbuka Ketahanan Pangan Provinsi Jawa Barat</p> <p>Peneliti:</p> <ol style="list-style-type: none"> <li>1. Adam Mukharil Bachtiar</li> <li>2. Dian Dharmayanti</li> <li>3. Hafizha Husnaisa</li> </ol> <p>Nama Jurnal: Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) Vol. 6, No.1, Maret 2017</p>	<p>Dalam penelitian ini, dilakukan pembuatan visualisasi data untuk dapat dilakukannya kegiatan pemantauan data tentang ketersediaan pangan melalui data yang disediakan oleh pemerintah pusat maupun daerah. Metode yang digunakan adalah Teknik Visualisasi Data Ben Fry dengan menggunakan metode regresi dan algoritma k-means dalam analisis algoritma.</p> <p>Hasil dari visualisasi ini berbentuk tabel dan grafik dalam menyajikan</p>

ISSN: 2089-9033	data sesuai kebutuhan yang. Visualisasi yang telah dirancang diimplementasikan dalam bentuk <i>software</i> .
-----------------	---

Berdasarkan tabel 2.1, dua penelitian yang dilakukan oleh Rayhanali Heiko Amier dan Johan Setiawan dan Intan Saktia Rahmawati dijadikan acuan dalam penelitian ini dan dalam melakukan pembuatan visualisasi data menggunakan *tools* Tableau dengan metode *Visual Data Mining* (VDM). Penelitian ketiga digunakan sebagai acuan dalam membuat visualisasi berdasarkan data COVID-19 dan Tingkat Pengangguran di Indonesia melalui data yang tersedia. Dari ketiga artikel jurnal diatas memiliki perbedaan dengan penelitian yang dilakukan adalah objek penelitian serta tidak dilakukannya prediksi dalam penelitian ini. Adapun dalam penelitian ini menggunakan metode yang sama dalam penelitian Rayhanali Heiko Amier dan Johan Setiawan dan Intan Saktia Rahmawati yaitu *Visual Data Mining* (VDM) serta *tools* yang digunakan yaitu Tableau.