

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Teori

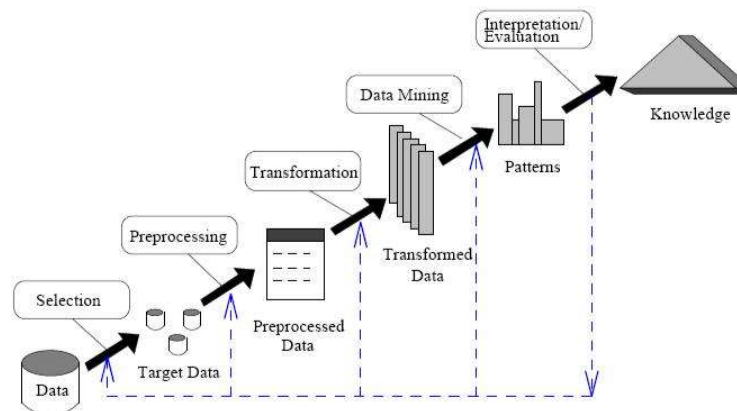
Penulis akan membahas beberapa teori yang akan menjadi pedoman dasar dalam melaksanakan penelitian, yaitu dengan menjelaskan mengenai pengertian dan istilah yang berkaitan dengan judul penelitian.

##### 2.1.1 Teori Umum

Penulis membahas beberapa teori umum yang berkaitan dengan pelaksanaan penelitian yang terdiri dari pengertian dan istilah secara umum.

##### 2.1.1.1 *Knowledge Discovery in Database (KDD)*

KDD merupakan sebuah aktifitas yang memiliki beberapa tujuan dan tahapan dalam mengumpulkan data yang berjumlah besar. Berikut tahapan dari *Knowledge Discovery in Database* secara detail (Arnol Holman Tua, 2020).



**Gambar 2.1** Tahapan *Knowledge Discovery in Database (KDD)*

Dari gambar 2.1 diatas maka tahapan KDD hingga menjadi suatu *Knowledge* atau pengetahuan baru adalah :

a. *Selection*

*Selection* digunakan untuk memilih variable yang akan diambil agar tidak adanya kesamaan dalam variable dan terjadinya perulangan dalam pengolahan data.

b. *Processing*

*Processing* terdapat 2 tahap, yaitu:

- a. Untuk menghilangkan data yang tidak diperlukan yaitu *Data Cleaning*.
- b. Data integration untuk mengidentifikasi data yang unik.

c. *Transformation*

*Transformation* adalah sebuah proses perubahan data, akan tetapi data tidak akan beda dari format existensi yang telah ditetapkan dengan prosedur pengolahan *datamining*.

d. *Datamining*

Proses utama dalam mendapatkan informasi dan data baru kemudian dilakukannya pemrosesan yang sesuai dengan prosedur yang sudah diberlakukan di *datamining*.

e. *Interpretation / Evaluation*

Untuk mengidentifikasi pola yang menarik kedalam data yang akan diidentifikasi.

*f. Knowledge*

Sebuah data yang telah didapat hasil untuk dijadikan sebuah informasi baru agar dapat digunakan dalam pengambilan sebuah keputusan.

Selain itu, KDD juga merupakan suatu proses dimana sebuah data yang mentah diolah menjadi sebuah informasi yang bermanfaat (Anita, M.Kom, Yulya, M., M.Kom, Sri, T., 2018).

### **2.1.1.2 Datamining**

*Datamining* dapat dikatakan sebagai metode pengolahan data yang berskala besar, sehingga sangat berperan penting dalam berbagai bidang (Rofiqo et al., 2018). Selain itu, *datamining* data digunakan dalam pengambilan keputusan. *Datamining* juga sering disebut dengan *Knowledge Discovery in Database* (KDD) atau *Pattern Recognition* (Sadewo et al., 2019). *Datamining* juga merupakan sebuah metode yang digunakan untuk pengolahan data yang berjumlah banyak (Maulida, 2018). Selain itu, *Datamining* juga digunakan dalam mengelompokkan penemuan pengetahuan di dalam sebuah basis data, serta *Datamining* merupakan sebuah proses yang menggunakan statistik, kecerdasan buatan, matematika, dan *machine learning* yang nantinya akan menghasilkan kelompok data yang bermanfaat (Elisa, 2018).

### 2.1.1.3 Clustering

Sebuah proses pembagian data ke beberapa kelompok dengan kesamaan data yang sama. Terdapat dua teknik metode *cluster* yaitu (Sadewo et al., 2019):

#### a. Hierarchical Clustering

Suatu metode pengelompokan data yang cara kerjanya menggunakan 2 data atau lebih yang memiliki kecenderungan mirip, kemudian dilanjutkan proses objek lain yang mempunyai kedekatan, proses terus berlangsung sampai cluster mencapai sebuah bentuk *tree*.

#### b. Non Hierarchical Clustering

Pada proses ini dimulai dengan memilih cluster yang diinginkan, kemudian proses *cluster* dilakukan tanpa harus mengikuti proses hirarki.

*Clustering* juga merupakan metode yang dalam prosesnya tidak tentu arah, yang berarti dalam metode ini tidak adanya training maupun pelatihan khusus dan tidak perlu adanya output (Suriani, 2020). Kelemahan dari metode *clustering* adalah metode tersebut dapat dikatakan sangat sensitif akan saat pemilihan *cluster* awal dan perhitungan dalam mencapai hasil yang optimal (Handoko, 2016).

### 2.1.1.4 K-Means

*K-Means* merupakan suatu metode pengelompokan data dengan *Non Hierarchical Clustering* yang dilakukan dengan memasukkan data ke dalam bentuk dua atau lebih kelompok (Annur, 2019). Langkah – langkah dalam melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut (Triyansyah & Fitriana, 2018) :

1. Tentukan jumlah *cluster*  $k$ .
2. Menentukan titik pusat awal dari setiap *cluster*.
3. Kelompokkan semua data/objek ke *cluster* terdekat dan untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (1)$$

**Rumus 2.1** Teori Jarak *Euclidean*

4. Hitung kembali pusat *cluster* dengan kelompok *cluster* yang sekarang. Pusat *cluster* adalah rata – rata dari semua data/objek dalam *cluster* tertentu.
5. Lakukan perhitungan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses clustering selesai. Dan jika adanya perbedaan maka akan kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

*K-Means* juga merupakan metode yang membagikan data menjadi beberapa kelompok yang memiliki karakteristik yang sama dan jika ada yang berbeda maka akan di masukkan kedalam kelompok lainnya (Bena Siti Ashari, Steven Christ Otniel, 2019).

#### **2.1.1.5 Rapidminer**

*Rapidminer* merupakan perangkat lunak yang berupa ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan

lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam penambangan teks, dan analisis prediktif (Dewi et al., 2019).

*Rapidminer* merupakan aplikasi *Open Source* dan dapat menjadi solusi dalam mengerjakan *datamining*, *text mining*, dan analisa prediksi. *Rapidminer* memiliki 500 operator *datamining*. *Rapidminer* juga merupakan aplikasi yang dapat berdiri sendiri dalam menganalisis sebuah data (Ardiansyah & Walim, 2018).

## **2.1.2 Teori Khusus**

### **2.1.2.1 UMKM**

UMKM merupakan sebuah usaha yang untuk mengembangkan perekonomian disetiap Negara, khususnya pada negera yang masih berkembang yang berdasarkan Undang – undang no. 20 pasal 1 Tahun 2008. Yang mengartikan UMKM berdasarkan usaha kecil, usaha menengah, dan usaha besar yang dimana memenuhi kriterial usaha berdasarkan undang – undang yang memiliki kekayaan bersih 50.000.000 sampai 10 milyar, sedangkan hasil penjualan mencapai 300 juta sampai 500 milyar per tahun (Sari & Santoso, 2019).

UMKM sudah menjadi peran yang sangat penting di Indonesia. Hal tersebut dikarenakan UMKM merupakan usaha yang berjumlah lebih banyak dibandingkan dengan usaha di bidang industri (Suci et al., 2017).

Selain itu, UMKM sudah menjadi perhatian bagi pemerintah untuk terus dikembangkan. UMKM juga sudah menjadi sector dalam menyediakan peluang besar bagi masyarakat yang ingin mencari pekerjaan dan terbukti dapat membantu masyarakat dalam mengatasi ekonomi yang krisis (Ilmaniati & Putro, 2019).

### **2.1.2.2 Internet**

Singkatan dari internet adalah *Interconnection Networking*. Internet merupakan sebuah jaringan yang dapat menghubungkan jaringan satu dengan lainnya agar komputer dapat melakukan komunikasi dengan komputer lainnya (Irawan, 2018).

Internet merupakan sebuah jaringan komputer yang sangat besar dan memiliki jutaan perangkat komputer yang menghubungkan jaringan satu ke jaringan komputer lainnya yang dapat digunakan untuk melakukan pertukaran informasi (Putri, 2020).

### **2.1.2.3 Media Sosial**

Sebelum adanya perkembangan dalam dunia teknologi, kita menonton dengan televisi yang hanya terdapat warna hitam putih. Dengan adanya perkembangan teknologi maka munculnya televisi digital yang dapat kita sambungkan ke internet. Selain itu, juga terdapat media sosial yang dulu kita gunakan media cetak dan sekarang kita sudah dapat menggunakan sambungan internet dan menggunakan berbagai macam media sosial yang semakin canggih dan mudah digunakan, seperti *Facebook*, *Instagram*, *Whatsapp*, dan sebagainya (Gani, 2020).

Media sosial merupakan komunikasi yang dilakukan melalui daring menggunakan sebuah situs *websites* yang mengubah komunikasi menjadi dialog interaktif. Dulu kita menggunakan media tradisional dengan media cetak, sekarang kita mengakses media sosial dengan internet (Wisti Aristika., 2019).

Media sosial merupakan media yang berbasis internet untuk digunakan dalam berbagai kegiatan seperti untuk melakukan komunikasi dengan pengguna lainnya dan untuk sebagai wawasan luas bagi pengguna dalam berinteraksi dengan pengguna media sosial lainnya (Budi Mahardhika & Sunariani, 2019).

Media sosial juga merupakan komunikasi alternatif yang sangat mudah digunakan oleh masyarakat, sehingga dapat mempersingkat waktu dalam komunikasi jarak jauh serta dapat dilakukan kapanpun dan dimana saja (Siti Nurul Intan & Hutabarat, 2020).

#### **2.1.2.4 Media online**

*E-commerce* saat ini sudah terbukti dapat membantu dalam kita mempromosikan usaha ataupun produk yang akan dijual. Dapat disebutkan bahwa proses jual dan pembelian sebuah produk secara internet atau *online* adalah *e-commerce*. *E-commerce* juga merupakan sebuah proses dalam membuat, mengelola sebuah hubungan komersial secara *online* (Arifianto & Coiri, 2018).



## 2.2 Kerangka Pemikiran

Kerangka pemikiran dari penelitian sebagai berikut :



**Gambar 2.2** Kerangka Pemikiran

Sumber : ( Penulis, 2020 )

## 2.3 Hipotesis Penelitian

Demikian penjelasan dari hipotesis yang menjadi atau dugaan penulis sementara terhadap penelitian, yaitu :

1. Diduga *datamining* dengan algoritma *K-means* dapat menentukan media sosial yang paling diminati oleh masyarakat di Kota Batam tepatnya pada kecamatan Batam Kota.
2. Diduga hasil pengolahan *datamining* dengan algoritma *K-means* dapat memberikan solusi bagi UMKM dalam menentukan strategi promosi di media sosial.

## 2.4 Penelitian Terdahulu

Penulis menemukan beberapa penelitian terdahulu yang dapat membantu dalam melakukan penelitian ini yang berkaitan dengan metode *datamining* dengan algoritma *clustering K-Means*.

1. Penelitian tentang “Penentuan Tingkat Minat Belanja Online Melalui Media Sosial Menggunakan Metode *Clustering K-Means*” (Anita, M.Kom, Yulya, M., M.Kom, Sri, T., 2018). Dari hasil penelitian tersebut menggunakan perhitungan *clustering K-means* dapat mengetahui media sosial yang paling sering digunakan dalam melakukan belanja online.
2. Penelitian tentang “Implementasi *Rapidminer* Dengan Metode *K-Means* (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)” (Agus Perdana, Riyani Wulan Sari, 2018). Dari hasil penelitian tersebut diharapkan pemerintah dapat lebih memperhatikan imunisasi campak pada balita khusus di provinsi yang keterbatasan dalam imunisasi campak pada balita.
3. Penelitian tentang “Penerapan *Datamining* Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode *K-Means Clustering* ( Studi Kasus di Program Studi TKJ Akademi Komunitas Solok Selatan)” (Handoko, 2016). Dari hasil penelitian tersebut dapat kita lihat akademik mahasiswa tersebut tergolong sedang karena IP rata – rata sekitar 2,50 s/d 3,00 dan jarak ke kampus terlampau jauh sekitar 5 M serta jumlah kehadiran berjumlah 85 s/d 90 kehadiran.
4. Penelitian tentang “Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode *Crisp –DM* dengan Algoritma *K-*

*Means Clustering*” (Astuti, 2019). Dari hasil penelitian adalah kita harus memiliki strategi dalam melakukan promosi berdasarkan omset masing – masing dalam meningkatkan penjualan.

5. Penelitian tentang “Penerapan *Clustering K-Means* Untuk Menentukan Pengaruh Media Sosial *Facebook* Terhadap Usaha Mikro , Kecil, dan Menengah (UMKM) Di Kecamatan Pekanbaru Kota” (Wisti Aristika., 2019). Dari hasil penelitian tersebut dapat membantu dalam menentukan seberapa besarnya pengaruh media sosial *Facebook* terhadap UMKM di Kecamatan Pekanbaru.
6. Penelitian tentang “Algoritma *K-Means* Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi” (Sadewo et al., 2019). Dari hasil penelitian ini dapat menjadi sebuah acuan bagi Pemerintah agar lebih memperhatikan dalam pembangunan lampu di jalan utama.
7. Penelitian tentang “Penerapan *Clustering* Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan *Datamining K-Means*” (Rofiqo et al., 2018). Dari hasil penelitian tersebut dapat diketahui terdapat 6 provinsi yang memiliki keluhan kesehatan yang tinggi, 18 provinsi dengan keluhan kesehatan yang sedang dan 9 provinsi yang memiliki keluhan kesehatan yang rendah.
8. Penelitian tentang “Penerapan *Datamining* Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan di Prov. DKI Jakarta Dengan *K-Means*” (Maulida, 2018). Dari hasil penelitian tersebut dapat menjadi perhatian pemerintah untuk lebih memperhatikan tempat wisata yang

perlu untuk di perbaiki ulang kembali agar pengunjung dapat berwisata dengan nyaman.

9. Penelitian tentang “Penerapan Metode *K-Means Cluster Analysis* Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa Internasional Class STMIK AMIKOM Yogyakarta” (Aranda & Natasya, 2016). Dari hasil penelitian tersebut bahwa 4 mahasiswa diarahkan untuk mengambil pemograman, 4 mahasiswa mengambil multimedia dan 3 – 5 mahasiswa mengambil jaringan komputer.
10. Penelitian tentang “Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan *Datamining* Algoritma *K-Means Clustering*” (Suriani, 2020). Dari hasil penelitian tersebut dapat diketahui bahwa masih belum efektif karena metode yang digunakan masih belum mendukung. Dan kategori dicari berdasarkan daerah yang sangat rawan, sedang, dan tidak rawan.
11. Penelitian tentang “Penerapan Teknik *Clustering* Sebagai Startegi Pemasaran Pada Penjualan Buku Di Tokopedia dan Shopee” (Rachman & Wasiyanti, 2019). Dari hasil penelitian tersebut dapat diketahui bahwa *clustering* dapat dilakukan untuk menganalisis pemasaran.
12. Penelitian tentang “Analisis *Datamining* Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing” (Triyansyah & Fitriana, 2018) . Dari hasil penelitian tersebut dapat diketahui bahwa terdapat 3 tipe sepatu yang paling diminati.

13. Penelitian tentang “*Analysis Of The K-Means Algorithm On Clean Water Customers Based On The Province*” (Windarto et al., 2019). To cluster information on the number of cleaned water clients by k-means strategy utilizing tools RapidMiner, the result of clustering with 3 clusters: C1: 6 areas: North Sumatera, South Sumatera, Bali, South Kalimantan, East Kalimantan and South Sulawesi, C2: 4 territories : DKI Jakarta, West Java, Central Java and East Java, C3: 24 territories: Aceh, West Sumatra, Riau, Jambi, Bengkulu, Lampung, Bangka Island, Riau Island, Yogyakarta, Banten, West Nusa Tenggara, Central Nusa Tenggara, West Kalimantan, Central Kalimantan, North Sulawesi, North Sulawesi, Central Sulawesi, Southwest Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua and Papua.
14. Penelitian tentang “*Using The K-Means Method For Diagnosing Cancer Stage Using The Pandas Library*”(Meniailov et al., 2019). The consider analyzed the measurable reliance between the factors that determine the condition of patients. It is appeared how it is conceivable to decide the patient’s having a place to a certain lesson on the premise of factual information - enlisted state variables. The created program arrangement, based on the Python Pandas library, makes it possible to classify on the premise of preparing tests, which guarantees a tall percentage of acknowledgment, as well as to distinguish the point of reference having a place to one of a few clusters. It ought to be famous that the depicted

*approach is widespread and can be utilized not only for biomedical frameworks, but too specialized, financial, etc.*

15. Penelitian tentang “*Implementation Data Mining With K-Means Algorithm For Clustering Distribution Rabies Case Area In Palembang City*” (Rahayu et al., 2020). *The result of information preparing utilizing RapidMiner program with the strategy of K-Means can be known from 16 sub-districts in Palembang, seven sub-districts counting the territorial cluster is very inclined to rabies (C0), whereas the four sub-districts included within the cluster Rabies-prone areas (C1), and five sub-districts included within the territorial cluster isn't inclined to rabies (C2). With the scourge bunch can be mediated to a program of Palembang wellbeing benefit that is fast-responsive and on target.*