

## BAB II

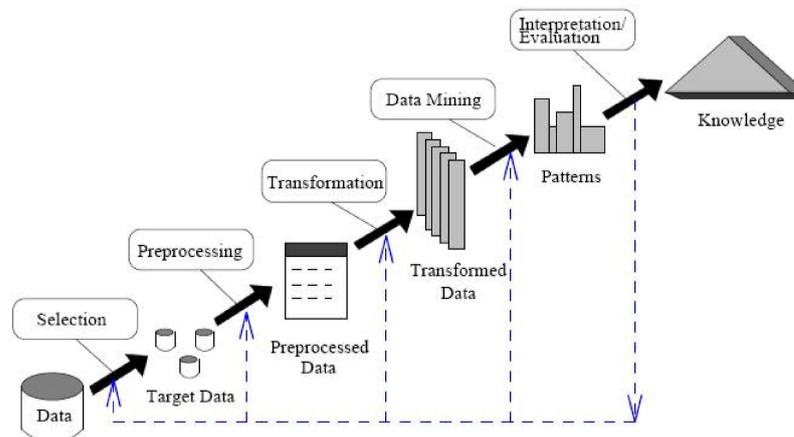
### TINJAUAN PUSTAKA

#### 2. 1. Teori Umum

Penulisan akan membahas beberapa teori yang akan menjadi pedoman dasar dalam melakukan penelitian, yaitu dengan menjelaskan mengenai pengertian dan istilah yang berkaitan dengan judul penelitian. Penulisan membahas beberapa teori umum yang berkaitan dengan pelaksanaan penelitian yang terdiri dari pengertian dan istilah secara umum.

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

KDD merupakan sebuah aktifitas yang memiliki beberapa tujuan dan tahapan dalam mengumpulkan data yang berukuran besar. Berikut adalah beberapa tahapan dari *Knowledge Discovery in Database* secara detail (Holman & Elisa, 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 pemilihan *atribut* yang akan dilakukan untuk melakukan pemosresan pada pengolahan *data mining*.

Sebelum melakukan tahap pencarian informasi dalam KDD dilakukan, sekumpulan data operasional perlu dilakukan dengan cara pemilihan data (Sundari et al., 2021).

*b. Processing*

*Processing* merupakan sebuah dasar untuk melakukan penghapusan data dan mengidentifikasi data menjadi unik.

Dilakukannya proses pembuangan duplikasi data, memeriksa data, dan memperbaiki kesalahan data sebelum dilakukannya proses *datamining* (Sundari et al., 2021).

*c. Transformation*

*Transformation* merupakan sebuah proses perubahan data, tetapi data yang diproses tidak akan berbeda dengan format existensi yang sudah ditetapkan prosedur pengelolaan *datamining*.

*d. Datamining*

*Datamining* merupakan sebuah proses dalam mendapatkan informasi dan data baru yang kemudian melakukan pemrosesan yang sesuai dengan prosedur yang sudah diberlakukan di *datamining*.

*Datamining* adalah sebuah pola atau informasi yang menarik data yang akan terpilih dapat dilakukan dengan teknik, metode, atau algoritma

secara bervariasi. Pemilihan metode sangat bergantung pada tujuan dan proses KDD (Sundari et al., 2021).

*e. Interpretation / Evaluation*

*Interpretation / Evaluation* merupakan pola informasi yang dihasilkan untuk ditampilkan dalam bentuk yang mudah untuk dimengerti.

Pada proses ini dilakukannya pemeriksaan pola dan informasi dalam bentuk yang mudah dipahami dan sesuai dengan fakta dan hipotesis sebelumnya (Sundari et al., 2021).

*f. Knowledge*

*knowledge* merupakan sebuah data yang telah didapatkan hasil untuk dijadikan sebuah informasi baru agar dapat digunakan dalam pengambilan sebuah keputusan.

### **2.1.2 *Datamining***

*Datamining* bisa dikatakan sebagai metode yang dilakukan untuk pengolahan data berskala besar, sehingga sangat berperan penting dalam berbagai bidang. Selain itu, data *datamining* digunakan untuk pengambilan keputusan. *Datamining* juga merupakan sebuah metode proses yang menggunakan statistik, kecerdasan buatan, matematika, dan *machine learning* yang akan menghasilkan kelompok yang bermanfaat.

### **2.1.3 *K-Means***

*K-means clustering* merupakan sebuah teknik yang biasanya digunakan untuk pencarian persamaan data ataupun kemiripan dan pengelompokan data antara beberapa data dengan data lainnya. *K-means* juga bisa diartikan sebagai

pengelompokkan data melalui *cluster* hingga pemberhentian yang sudah ditetapkan. Dari beberapa data kelompok yang mempunyai ciri khas berbeda dengan data yang ada ke dalam kelompok yang berbeda dilakukannya pegabungan data menggunakan metode *K-means* (Jtik et al., 2022).

#### **2.1.4 Clustering**

Kegiatan pengelompokkan berbagai macam pengetahuan atau pengamatan dalam melakukan pendekatan dari suatu objek ke objek lainnya yang mempunyai kesamaan disebut *Clustering*. *Cluster* mempunyai beberapa deretan baris pengetahuan yang mempunyai kesamaan dan perbedaan terhadap *cluster* lainnya (Purwanto et al., 2022).

#### **2.1.5 SPSS**

SPSS adalah suatu singkatan dari *Statistical Product and Service Solution*. SPSS merupakan bagian dari proses analisa, dalam penyediaan akses data. SPSS juga bisa membaca jenis-jenis data ataupun memasukkan data dengan jumlah yang banyak langsung ke dalam SPSS Data Editor (Janna & Herianto, 2021).

SPSS dapat membaca dan menulis data file teks ASCII, dan juga dapat ke tabel database eksternal melalui ODBC dan SQL. Awal versi SPSS merupakan versi yang dirancang untuk melakukan proses *batch* di *mainframe*, seperti IBM dan versi ICL, yang pengawalannya menggunakan kartu untuk input.

Program yang ada di SPSS tingkatan adalah : Analisis struktur moment yang memungkinkan pemodelan persamaan terhadap struktual dan struktur analisis jalur, dan mempunyai kemampuan yang mendasar seperti analisis regresi linier, ANOVA, ataupun ANCOVA (Bhirawa, 2020).

## **2.1.6 Uji Validitas dan Reabilitas**

### **2.1.6.1 Uji Validitas**

Uji validitas adalah pengujian yang sejauh mana alat pengukur dapat mengukur. Alat ukur yang dinyatakan akurat apabila validitas instrument semakin tinggi dalam suatu pengukuran data. Pada pengujian ini sangat penting dilakukan agar tidak terjadinya penghasilan data yang menyimpang dari gambaran variabel yang dimaksud dari pertanyaan (Amanda et al., 2019).

Awal pengambilan data keputusan dinyatakan valid bila korelasi rhitung  $> 0,099$  (rtabel) maka dapat disimpulkan bahwa data tersebut valid, tetapi jika korelasi rhitung  $< 0,099$ (rtabel) maka data tersebut disimpulkan tidak valid (Ilmi, 2021).

### **2.1.6.2 Uji Reliabilitas**

Uji reliabilitas adalah pengujian instrumen untuk mengetahui apakah instrument tersebut dapat digunakan untuk pengumpulan data atau tidak dapat mengumpulkan data. Dalam pengujian reliabilitas dilakukan rumus alpha.

Pengambilan data keputusan dinyatakan realibilitas jika Cronbach's Alpha  $> 0,60$  maka dapat disimpulkan bahwa data tersebut reliabel, tetapi jika Cronbach's Alpha  $< 0,60$  maka data tersebut disimpulkan tidak reliabel (Ilmi, 2021).

## **2.1.4 Uji Asumsi Klasik**

### **2.1.4.1 Uji Normalitas**

Uji normalitas merupakan pengujian untuk melakukan proses regresi, variabel independen, variabel dependen, ataupun variabel keduanya untuk mengetahui apakah data yang diuji normal atau tidak.

Pada uji normalitas diuji dengan tingkat signifikansi 0,05 atau 50%. Jika hasil uji  $> 0,05$  maka dapat disimpulkan bahwa data tersebut normal, tetapi jika hasil uji  $< 0,05$  maka data tersebut disimpulkan tidak normal (Ilmi, 2021).

#### **2.1.4.2 Uji Multikolienaritas**

Uji multikolienaritas adalah hubungan linier yang mendekati antara variabel independen pada model regresi dan linier sempurna. Tujuan dari uji multikolienaritas adalah untuk menguji model regresi apakah ditemukannya korelasi antara variabel bebas (Tobuhu et al., 2022).

Untuk melakukan uji multikolienaritas dapat dilihat pada nilai *Variant Inflation Factor* (VIF). Jika dari nilai VIF  $> 10$  maka terlihat adanya multikolienaritas dan jika VIF  $< 10$  maka tidak akan terlihat adanya multikolienaritas (Ilmi, 2021).

#### **2.1.4.3 Uji Heteroskedastisitas**

Uji heteroskedastisitas mempunyai tujuan untuk mengetahui apakah sebuah regresi mempunyai ketidaksamaan terhadap variasi yang konstan dari satu pengamatan ke pengamatan lainnya. Tidak terjadinya heteroskedastisitas merupakan model regresi yang baik (Tobuhu et al., 2022).

Dalam pengamatan ini, dilakukannya pengujian menggunakan uji glejser yang merupakan uji hipotesis yang bertujuan untuk mengetahui apakah model regresi mempunyai indikasi heteroskedastisitas.

Pengambilan keputusan uji glejser dapat dilihat jika signifikansi hasil uji  $> 0,05$  maka tidak terjadi heteroskedastisitas dan jika hasil uji  $< 0,05$  maka terjadi heteroskedastisitas.

### **2.1.5 Regresi Linier Sederhana**

Analisis regresi merupakan studi yang bergantung dari suatu variabel yang disebut tidak bebas (*dependent variable*), pada satu ataupun lebih variabel, yang dimana variabel tersebut bertujuan untuk memperkirakan nilai dari variabel yang menjelaskan sudah diketahui. Variabel yang menjelaskan sering adalah variabel bebas (*independent variable*) (Bhirawa, 2020).

Menurut (Trianggana, 2020), Regresi Linear Sederhana adalah sebuah metode statistik yang digunakan untuk menguji sejauh mana hubungan atau pengaruh antara variabel X terhadap variabel Y.

Untuk melakukan uji signifikansi, maka menggunakan uji-t. Bentuk hasil uji-t mempunyai pengujian dua arah yang dimana jika,  $t \text{ hitung} > t \text{ tabel}$  maka  $H_0$  akan ditolak dan  $H_1$  diterima. Jika  $t \text{ hitung} < t \text{ tabel}$  maka  $H_0$  diterima dan  $H_1$  ditolak.

### **2.1.6 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.

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.

## **2. 2. Teori Khusus**

### **2.2.1 Transportasi**

Transportasi merupakan sebuah usaha dan kegiatan untuk mengangkut atau membawa barang atau penumpang dari suatu tempat ke tempat yang lainnya. Untuk setiap bentuk transportasi terdapat empat unsur pokok transportasi, yaitu: jalan, kendaraan dan alat angkutan, tenaga penggerak, dan terminal. Pengangkutan atau pemindahan barang atau penumpang dengan transportasi merupakan pencapaian tempat tujuan dan menciptakan atau menaikkan utilitas dan kegunaan dari yang diangkut.

### **2.2.2 Transportasi Online**

Transportasi online merupakan pelayanan jasa transportasi yang berbasis online dalam setiap kegiatan transaksinya, mulai dari pemesanan, pemantauan jalur, pembayaran dan penilaian terhadap pelayanan jasa tersebut. Transportasi online adalah salah satu bentuk dari penyelenggara lalu lintas dan angkutan jalan yang berjalan dengan mengikuti serta memanfaatkan perkembangan ilmu pengetahuan (teknologi).

Transportasi online merupakan transportasi yang berbasis suatu aplikasi tertentu, dimana konsumen memesan suatu sarana transportasi melalui system aplikasi di dalam smartphone. Saat konsumen melakukan pemesanan dengan

menggunakan aplikasi, detail pemesanan seperti jarak tempuh, harga, identitas pengemudi, lama waktu pengemudi tiba ke lokasi konsumen. Seluruh identitas pengemudi sudah diketahui secara pasti karena perusahaan pengelola telah melakukan proses verifikasi terlebih dahulu sebelum melakukan Kerjasama kemitraan dengan pengemudi.

### 2.2.3 Gojek

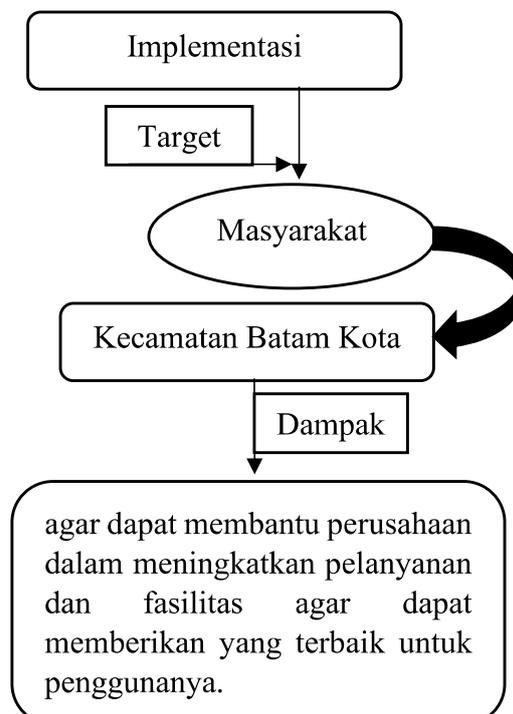
Gojek adalah sebuah perusahaan yang berjiwa sosial dalam memimpin revolusi industri transportasi ojek. Gojek dengan para pengemudi ojek yang sudah berpengalaman dapat dijadikan solusi utama untuk melakukan pengiriman barang, pemesanan antar makanan, dan belanjaan.

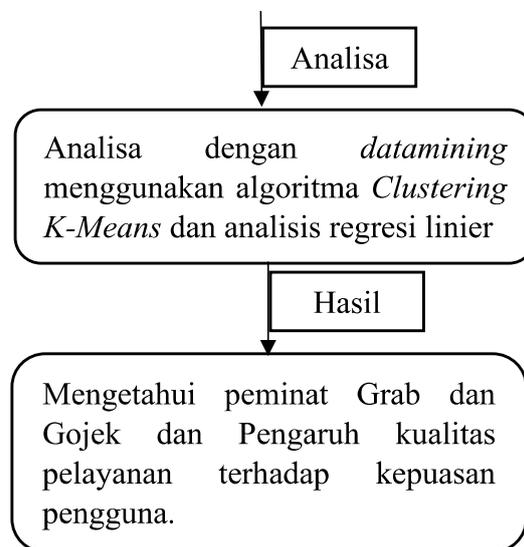
### 2.2.4 Grab

Grab adalah perusahaan teknologi yang menghubungkan penumpang dengan pengemudi taksi melalui aplikasi online.

## 2.3. Kerangka Pemikiran

Kerangka pemikiran dari penelitian sebagai berikut:





**Gambar 2. 2** Kerangka Pemikiran

**Sumber :** (penulis, 2022)

#### **2. 4. Hipotesis Penelitian**

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

1. Diduga kualitas pelayanan berpengaruh terhadap kepuasan pengguna Gojek dan Grab berdasarkan data dari masyarakat di Kota Batam tepatnya pada Kecamatan Batam Kota.
2. Diduga dapat memberikan solusi bagi masyarakat dalam menentukan kepuasan kualitas pelayanan terhadap pengguna Gojek dan Grab.

#### **2. 5. Penelitian Terdahulu**

Penulis menemukan beberapa penelitian terdahulu yang dapat membantu penulis dalam melakukan penelitian yang berkaitan dengan metode *datamining* dengan algoritma *clustering K-Means* dan analisis regresi linier.

1. Penelitian tentang “Analisis Pengaruh Kualitas Pelayanan dan Kualitas Produk Terhadap Repurchase Intention Konsumen Skincare Scarlett (Studi pada E-Commerce Shopee)” (Ilmi, 2021). Dari hasil penelitian tersebut terdapat penjelasan bahwa Kualitas Pelayanan (X1) tidak berpengaruh secara signifikan terhadap Minat Beli Ulang (Y). Variabel Kualitas Produk (X2) berpengaruh secara signifikan terhadap Minat Beli Ulang (Y). penelitian ini menunjukkan bahwa Kualitas Pelayanan (X1) dan Kualitas Produk (X2) berpengaruh secara signifikan terhadap Minat Beli Ulang (Y). Pada penelitian ini menggunakan metode analisis regresi linier berganda yang menggunakan SPSS 22.
2. Penelitian tentang “Analisis Faktor Yang Mempengaruhi Penumpang Angkutan Umum Beralih ke Transportasi Online Go-Jek menggunakan Metode *K-Means Clustering*” (Jtik et al., 2022). Dari hasil penelitian tersebut terdapat penjelasan bahwa umur mempunyai nilai paling tinggi untuk faktor peralihan jangka tempuh angkutan umum yaitu di posisi ketiga, faktor peralihan jarak pada Go-Jek berada di posisi kelima, faktor peralihan tarif berada di posisi keempat dan faktor peralihan mudah di dapat berada di posisi kedua.
3. Penelitian tentang “Perbandingan Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah Menggunakan Algoritma K-Means dan K-Medoids” (Rofik, 2021). Berdasarkan hasil perhitungan yang diperoleh K-Means adalah 3 iterasi siswa yang merasa puas dengan pelayanan sekolah dengan jumlah siswa 276, sedangkan K-Medoids lebih baik karena hanya memperoleh 2

iterasi dengan hasil siswa yang merasa puas dengan pelayanan sekolah dengan jumlah siswa 324 siswa, sehingga algoritma K-Medoids lebih cocok digunakan dalam mengklaster kepuasan siswa disekolah.

4. Penelitian tentang “Mengukur Tingkat Kepuasan Masyarakat Pada Pelayanan Kepolisian Resor(Polres) Dumai Menggunakan Algoritma K-Means Clustering” (Burelia et al., 2022). Pada pengukuran tingkat kepuasan masyarakat terhadap Polres Dumai menggunakan K-Means Clustering dapat mengukur kepuasan yang bisa menjadi acuan pada pelayanan Polres Dumai dalam meningkatkan mutu pada pelayanan Polres terhadap masyarakat Dumai.
5. Penelitian tentang “Penerapan Data Mining Untuk Menganalisa Tingkat Kepuasan Pelanggan Telkomsel Terhadap Sikap Pelayanan Caroline Officer Dengan Metode Menggunakan Algoritma K-Means Clustering” (Elisna et al., 2019). Untuk menampilkan hasil kepuasan pelanggan dengan tepat dilakukannya proses dengan menentukan sampel data pada pelanggan menggunakan K-Means Clustering.
6. Penelitian tentang “Data Mining Kepuasan Pelanggan Terhadap Pelayanan PDAM Menggunakan Metode Clustering” (Simanjuntak et al., 2022). Dari hasil perhitungan pada pelayanan PDAM, indikator tanggung jawab memiliki nilai tertinggi dibandingkan indikator lainnya.
7. Penelitian tentang “Clustering Kepuasan Layanan Pengguna Bus Trans Binjai Dengan Metode Cluster Data Mining Studi Kasus Dinas Perhubungan Kota Binjai” (Sundari et al., 2021). Berdasarkan hasil perhitungan dengan metode

clustering dapat disimpulkan bahwa centroid 1, centroid 2 dan centroid 3 mendapatkan penilaian 4(puas) dari Pengajar/Guru, Ibu Rumah Tangga, dan Pelajar/Mahasiswa.

8. Penelitian tentang “Pengaruh Kualitas Pelayanan Transportasi Online Gojek Terhadap Kepuasan Pelanggan Pada Mahasiswa/I Administrasi Niaga Politeknik Negeri Jakarta” (Rifaldi et al., 2019). Dari hasil analisis regresi linier sederhana terdapat perhitungan  $Y = 31,786 + 0,672X$  yang artinya jika pengaruh kualitas pelayanan (X) sama dengan nol, maka Kepuasan pelanggan (Y) sebesar 31,786.
9. Penelitian tentang “Implementasi Algoritma K-Means Untuk Clustering Sentimen Pada Opini Kualitas Pelayanan Jasa Penerbangan” (Safitri et al., 2021). Hasil dari persentase yang diperoleh dari perhitungan algoritma K-Means yaitu pada maskapai Garuda Indonesia sebesar 71,8%, Air Asia sebesar 52,8% dan Lion Air sebesar 62,1%.
10. Penelitian tentang “Implementasi Data Mining Clustering Tingkat Kepuasan Konsumen Terhadap Pelayanan Go-Jek” (Sinaga et al., 2021). Berdasarkan dari hasil penelitian konsumen Gojek di Kota Pematangsiantar yang merasa puas terhadap pelayanan Go-Jek sebanyak 43 konsumen dan sebanyak 77 konsumen merasa tidak puas akan pelayanan Go-Jek.
11. Penelitian tentang “Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan” (Pradana et al., 2020). Dari data pelanggan 2.000 data terdapat hasil kluster sebanyak 2 dengan kluster 0 sebanyak 1303 atau sebesar 65% dan kluster 2 sebanyak 697 atau sebesar 35%.

12. Penelitian tentang “Analisis Cluster Kepuasan Pengguna Terhadap Layanan Shopee Menggunakan Algoritma K-Means” (Patimah et al., 2021). Dari hasil penelitian dilakukan pengujian menggunakan metode Black Box. Berdasarkan hasil UAT pada aspek desain, kemudahan pengguna, kegunaan dan fungsional yang memperoleh hasil penerimaan pengguna sebesar 82,86% (Sangat Baik).
13. Penelitian tentang *“Implementation Of Clustering Unsupervised Learning Using K-Means Mapping Techniques”* (Abbas et al., 2021). *Based the study result, clustering on the proportion of workers in the manufacturing industry sector using a K-Means mapping techniques apply by using 2 cluster. Of the 24 manufacturing industry sectors, only about 21% have a large workforce absorption.*
14. Penelitian tentang *“Using Big Data Fuzzy K-Means Clustering and Information Fusion Algorithm In English Teaching Ability Evaluation”* (Zhen, 2021). *The feature information and combines the big data information fusion and K-Means Clustering Algorithm to realize the clustering and integration of the index parameters of English teaching ability. Research has shown that the method in this paper is more accurate in evaluating English teaching ability and improves the efficiency of the use of English teaching resources.*
15. Penelitian tentang *“Design of English Learning Effectiveness Evaluation System Based on K-Means Clustering Algorithm”* (Zhang, 2021). *Experimental results show that the system can effectively improve the*

*shortcomings of the traditional system and realize the high-quality evaluation of student's English learning effectiveness.*