BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Data yang digunakan pada penelitian ini adalah data pengadaan barang umum tahun 2019-2021 sebanyak 4.067 baris data yang diambil di PT Percetakan dengan izin dari kepala seksi lansokmansak. Penelitian ini dilakukan dengan mengacu pada kerangka kerja *Knowledge Discovery in Databases (KDD)* yang sudah dijelaskan pada Bab 2 (Nofriansyah & Mariami, 2021).

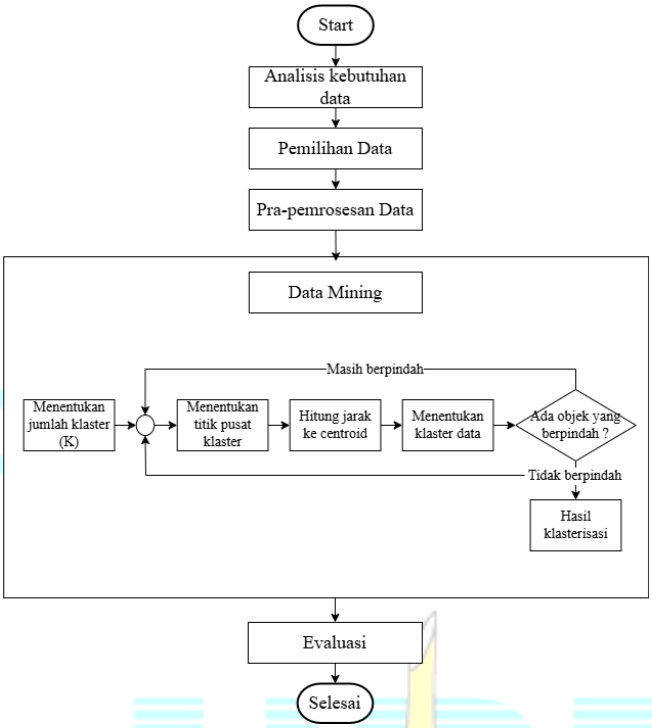
3.2 Proses Bisnis Perusahaan dan Kepentingan Supplier

Berikut merupakan penjelasan Proses Bisnis Pengadaan di PT Percetakan (I & MZ HAprilyanti S, 2018):

1. Identifikasi Kebutuhan: Perusahaan mengidentifikasi kebutuhan pengadaan barang atau jasa berdasarkan permintaan internal.
2. Pencarian *Supplier*: Dilakukan riset untuk mencari *supplier* yang memenuhi kriteria berdasarkan rekam jejak sebelumnya maupun referensi.
3. Penawaran Harga: *Supplier* yang terpilih diminta untuk memberikan penawaran harga dan spesifikasi barang atau jasa yang ditawarkan.
4. Proses Pemesanan: Dilakukan pembuatan *Purchase Order (PO)* dan persetujuan kontrak untuk pengadaan barang atau jasa.
5. Penerimaan Barang: Barang atau jasa yang telah dipesan dikirim oleh *supplier* dan diperiksa kesesuaiannya dengan spesifikasi yang telah disepakati.
6. Evaluasi *Supplier*: Perusahaan melakukan evaluasi terhadap *supplier* berdasarkan kinerja.

3.3 Prosedur Penelitian

Prosedur penelitian ini terdiri dari beberapa tahap yang sistematis, mulai dari menganalisis kebutuhan data, pemilihan data hingga analisis hasil klasterisasi. Tahapan prosedur penelitian dapat dilihat pada Gambar 3.1 (Rafi, 2020).



Gambar 3. 1 Prosedur Penelitian

1. Tahapan Prosedur Penelitian

1. Analisis Kebutuhan Data

Supplier yang dianalisis dalam penelitian ini adalah *supplier* yang berkontribusi dalam pengadaan barang umum perusahaan, seperti alat tulis kantor (ATK), perlengkapan produksi non-bahan utama, dan kebutuhan logistik lainnya. Berikut variabel utama yang akan digunakan dalam penelitian (Rafi, 2020):

1. *Supplier* : Identitas unik dari setiap *supplier*.
2. Jumlah Pesanan : Total kuantitas barang yang dipesan oleh perusahaan.
3. Harga Satuan : Biaya per unit barang.
4. Rentang Pengiriman : Waktu antara pemesanan barang hingga diterimanya barang oleh perusahaan.

2. Pemilihan Data

Pemilihan data dilakukan untuk memastikan bahwa data yang digunakan relevan dengan tujuan penelitian, yaitu mengelompokkan *supplier* berdasarkan kinerja menggunakan algoritma *K-Means*. Beberapa variabel yang digunakan dalam penelitian ini meliputi *Supplier*, Jumlah Pesanan, Harga Satuan dan Rentang Pengiriman. Variabel ini dipilih karena dianggap memiliki keterkaitan langsung dalam menilai kinerja *supplier*. Data yang tidak relevan,

duplikat, atau memiliki nilai kosong dihapus untuk menjaga kualitas analisis. Selain itu, *outlier* pada data diperiksa untuk menghindari distorsi pada hasil klasterisasi (Rafi, 2020). Berikut data yang sudah dipilih pada tabel 2.

Supplier	Jumlah Pesanan	Harga Satuan	Rentang Pengiriman
CPTSMM	822	11000000	45,55555556
CV. BT	200	16200000	91
CV. CM	125	31000000	30
CV. JMA	258	57500000	40,91358025
CV. KC	668	9000000	90
CV. LM	171	17500000	35,28571429
CV. MAB	467	63000000	31,28571429
CV. SAS	115	32000000	33,09090909
CV. SS	324	39000000	55,11111111
…	…	…	…
PT. WPO	109	16250000	39
PT. YAT	125	16800000	38,37692308

Tabel 1 Data yang Digunakan

3. Pra-pemrosesan Data

Pada tahap ini dilakukan pembersihan, transformasi dan normalisasi data untuk menghilangkan data yang duplikat dan tidak relevan untuk menangani *missing value* dan memperbaiki kesalahan *input* dengan *Google Colab* menggunakan *Phyton*. Metode normalisasi yang digunakan dalam analisis ini adalah *Min-Max Scaling* dan *Z-score Normalization* (Luh et al., 2022).

a. Min-Max Scaling

Min-Max Scaling digunakan untuk merubah nilai data ke dalam rentang [0,1] agar semua atribut memiliki skala yang sama (Luh et al., 2022). Rumus *Min-Max Scaling* adalah sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' = Nilai hasil normalisasi

X = Nilai asli

X_{min} = Nilai minimum dalam dataset

X_{max} = Nilai maksimum dalam dataset

Dengan metode ini, nilai terkecil dalam dataset akan menjadi 0, sedangkan nilai terbesar akan menjadi 1 (Luh et al., 2022).

b. Z-score Normalization

Z-score Normalization digunakan untuk merubah data sehingga memiliki distribusi dengan *mean* = 0 dan standar deviasi = 1 (Luh et al., 2022). Rumusnya adalah:

$$Z = \frac{X - \mu}{\sigma}$$

Keterangan**:**

Z = Nilai hasil normalisasi

X = Nilai asli

μ = Rata-rata (*mean*) dari dataset

σ = Standar deviasi dari dataset

Metode ini digunakan jika data memiliki distribusi yang berbeda-beda dan perlu dinormalisasi agar memiliki distribusi normal (Luh et al., 2022).

4. *Data Mining*

Tahapan ini adalah tahapan yang menerapkan metode dari *data mining* untuk mengolah data yang ada. Algoritma yang digunakan adalah algoritma *K-means* untuk mengelompokkan *supplier* berdasarkan kinerja. Dalam menentukan jumlah klaster optimal dengan *Elbow Method* dan *Silhouette Score* (Hananto et al., 2021).

1. Menentukan nilai K optimal

a. Metode *Elbow*

Metode ini menggunakan *Within-Cluster Sum of Squares (WCSS)*, yaitu total jarak kuadrat antara setiap titik data ke *centroid* klasternya. Rumus *WCSS* adalah:

$$WCSS = \sum_{i=1}^{K} \sum_{x \in C_i} || \ x - \mu_i \ ||^2$$

Keterangan:

x = Data dalam klaster

μ_i = Centroid klaster ke-i

C_i = Klaster ke-i

K = Jumlah klaster

Setelah menghitung *WCSS* untuk beberapa nilai K, hasilnya diplot dalam grafik. Titik siku (*elbow*) dari grafik menunjukkan K yang optimal (Hananto et al., 2021).

b. Metode *Silhouette Score*

Metode ini mengukur seberapa baik setiap data berada dalam klasternya dan seberapa jauh dari klaster lain. Rumusnya adalah sebagai berikut:

$$S = \frac{b - a}{\max(a, b)}$$

Keterangan:

S = Skor *Silhouette*

a = Rata-rata jarak data ke titik lain dalam klaster yang sama

b = Rata-rata jarak data ke titik terdekat di klaster lain

Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana semakin mendekati 1, semakin baik klasterisasi (Hananto et al., 2021).

2. Mengulang Proses *K-Means* (*Re-inisialisasi Centroid*)

Di tahap ini jika hasil klasterisasi masih kurang optimal, bisa mengulang menggunakan *K-Means++* untuk perulangannya, alasannya untuk menghasilkan hasil lebih stabil dan menghindari *centroid* awal yang buruk (Hananto et al., 2021).

5. Evaluasi Klaster

Setelah menentukan K dan menjalankan *K-Means*, hasil klasterisasi perlu dievaluasi, metode yang digunakan adalah *Davies-Bouldin Index (DBI)* untuk mengukur seberapa baik klaster terpisah dan seberapa kompak klaster tersebut. Semakin rendah nilai *DBI* maka klaster lebih baik (Hananto et al., 2021).

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d_{ij}} \right)$$

Setelah proses klasterisasi selesai, analisis hasil klasterisasi akan menggunakan *scatter plot* untuk visualisasi. Berdasarkan pengolahan data yang dilakukan dengan 4 variabel yang ada,

nantinya akan menghasilkan 3 *cluster* yaitu klaster dengan kinerja tinggi, kinerja sedang dan kinerja rendah (Sari, 2018).

Hasil Klasterisasi yang didapat nantinya akan dilakukan tindakan. Klaster tinggi akan direkomendasikan sebagai mitra kerja sama jangka panjang. Klaster sedang akan dilakukan pembinaan untuk peningkatan layanan terhadap perusahaan. Klaster rendah akan diberikan evaluasi terkait efektivitas kerja sama (I & MZ HAprilyanti S, 2018).