

SKRIPSI

PENGELOMPOKAN MINAT BACA MASYARAKAT KABUPATEN MAGELANG
MENGUNAKAN METODE
KMEANS CLUSTERING
(STUDI KASUS : DINAS PERPUSTAKAAN DAN KEARSIPAN KABUPATEN
MAGELANG)



ACHMAT MUJAFAR

NPM. 17.0504.0085

PROGRAM STUDI TEKNIK INFORMATIKA
S1 FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MAGELANG
FEBRUARI , 2022

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Menurut Turban (2005), Perkembangan teknologi informasi saat ini menjadikan suatu informasi sebagai elemen yang penting dalam perkembangan masyarakat. Penyajian informasi tidak sepadan dengan kebutuhan informasi yang sangat tinggi, sehingga informasi tersebut perlu digali lebih dalam dari data yang jumlahnya besar. Penggalan suatu informasi atau pola yang penting atau menarik dari data dalam jumlah besar digunakan para pengambil keputusan dalam memanfaatkan gudang data. Proses penggalan ini menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar disebut juga sebagai *data mining*.

Selama ini permasalahan penggalan informasi di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang masih tergolong rendah. Hal ini dibuktikan dengan banyaknya kategori buku yang berada di rak-rak buku perpustakaan. Untuk meningkatkan minat baca dilakukan kebijakan yang diperlukan untuk menentukan minat baca tersebut. Maka petugas perpustakaan bisa mengetahui kategori buku yang menjadi minat baca masyarakat. Sehingga data yang digunakan adalah data transaksi buku.

Dalam hal ini data transaksi buku dapat dijadikan sebagai basis data untuk dasar kebijakan pengelolaan data perpustakaan yang tepat dan akurat. Sehingga dengan adanya *data mining* dengan menggunakan data *cluster Algoritma K-Means* memudahkan petugas perpustakaan dalam menentukan minat baca masyarakat atau pengunjung perpustakaan tersebut.

Alasan dilakukannya penerapan metode pengelompokan *K-Means* ini adalah untuk melakukan pengelompokan minat baca masyarakat berdasarkan jumlah peminjaman buku pada Dinas Perpustakaan dan Kearsipan Kabupaten Magelang sehingga menuangkan kedalam bentuk

Skripsi yang berjudul “Pengelompokan Minat Baca Masyarakat Kabupaten Magelang Menggunakan Metode *K-Means Clustering* (Studi Kasus : Dinas Perpustakaan dan Kearsipan Kabupaten Magelang)”..

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka masalah penelitian ini yaitu apakah metode *K-Means Clustering* berdasarkan jumlah peminjaman berdasarkan kategori buku bisa memberikan gambaran minat baca masyarakat di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang ?.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diuraikan, penelitian ini bertujuan untuk :

1. Mengelompokkan minat baca masyarakat sesuai jumlah peminjaman buku di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang menggunakan metode *K-Means Clustering*.
2. Mengetahui kategori buku yang paling banyak diminati oleh masyarakat di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang.

1.4 Manfaat Penelitian

Manfaat penelitian yang diharapkan apabila tujuan penelitian tercapai adalah:

1. Mengetahui seberapa besar jumlah peminjaman buku.
2. Perpustakaan dapat mengetahui buku yang diminati di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang.

BAB 2

TINJAUAN PUSTAKA

2.1 Penelitian Relevan

Penelitian yang dilakukan Widya Safira Aziz dan Deni Atmajaya (2016) dalam Jurnal Ilmiah ILKOM Volume 8 Nomor 2 melakukan penelitian dengan judul “Pengelompokan Minat Baca Mahasiswa Menggunakan *Metode K-Means*”, menyatakan bahwa permasalahan yang dihadapi yaitu penerapan sistem yang diterapkan di Perpustakaan Utsman Bin Affan Universitas Muslim Indonesia yang berjalan kurang efektif. Sehingga cara yang digunakan untuk mengelola data tersebut secara komputerisasi. Salah satu cara mengelola data tersebut yaitu dengan memanfaatkan metode *K-means*, setiap data buku akan dikelompokkan menjadi 3 *cluster*, yaitu prioritas tinggi, dipertimbangkan dan bukan prioritas untuk perencanaan penambahan koleksi buku.

Penelitian yang dilakukan Agus Perdana Windarto (2017) dalam *International Journal Of Artificial Intelligence Research* Vol 1, No 2 December 2017 melakukan penelitian dengan judul “*Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method* “, menyatakan bahwa Produksi beras Indonesia memang begitu tinggi sehingga tidak dapat memenuhi kebutuhan penduduknya, akibatnya Indonesia masih harus mengimpor beras dari penghasil pangan lainnya negara. Salah satu penyebab utamanya adalah populasi yang sangat besar. Statistik menunjukkan bahwa di kisaran 230-237 juta orang, makanan pokok makanan semua warga adalah nasi jadi jelas kebutuhannya akan nasi menjadi sangat besar. Sehingga Metode *clustering* yang digunakan dalam penelitian ini adalah Metode *K-Means*. *Data Centroid* untuk klaster tingkat impor tinggi 7429180 dan 2735452,25, *Data Centroid* untuk tingkat impor sedang *cluster* 1046359.5 dan 337703.05 serta *Data Centeroid* untuk impor rendah *cluster* level 185559.425 dan 53089.225. Hasilnya adalah penilaian berdasarkan indeks impor beras dengan 2 negara cluster impor tinggi yaitu Vietnam dan Thailand, 4 cluster tingkat sedang negara importir yaitu China, India, Pakistan dan 4 Lainnya serta rendah negara *cluster* yang diimpor yaitu USA, Taiwan, Singapore dan Myanmar. 6

Penelitian yang dilakukan Windania Purba, Saut Tamba, Jepronel Saragih (2018) dalam *IOP Conf. Series: Journal of Physics: Conf. Series* melakukan penelitian dengan judul “*The Effect Of Mining Data K-Means Clustering Toward Students Profile Model Drop Out Potential*”. menyatakan bahwa permasalahan yang dihadapi yaitu Tingkat keberhasilan siswa yang tinggi dan rendahnya kegagalan siswa dapat mencerminkan kualitas suatu perguruan tinggi. Salah satu faktor siswa gagal adalah putus sekolah.. Sehingga cara yang digunakan untuk mengatasi masalah tersebut yaitu *Data Mining* dengan *K-means Clustering* dapat diterapkan. Metode *K-Means Clustering* akan diimplementasikan untuk mengelompokkan siswa yang berpotensi drop out. Pertama data hasil akan di clustering untuk mendapatkan informasi kondisi seluruh siswa. Berdasarkan model yang diambil ditemukan bahwa siswa berpotensi drop out karena siswa kurang bersemangat dalam belajar, orang tua tidak didukung, siswa malu-malu dan waktu berperilaku siswa yang kurang. Hasil dari proses *K-Means Clustering* dapat diketahui bahwa mahasiswa yang lebih berpotensi drop out berada pada Cluster 1 menyebabkan *Credit Total System*, *Quality Total*, dan *Grade Point Average (IPK)* terendah dibandingkan antara cluster 2 dan 3.

Penelitian yang dilakukan Yuda Irawan (2019) yang dipublikasikan pada *Journal of Applied Engineering and Technological Science Vol 1(1) 2019: 17-29* Melakukan Penelitian Dengan Judul “*Implementation Of Data Mining For Determining Majors Using K-Means Algorithm In Students Of SMA Negeri 1 Pangkalan Kerinci*”. menyatakan bahwa permasalahan yang dihadapi yaitu pemilihan jurusan yang tidak tepat sehingga bisa merugikan mahasiswa untuk minat dan karir mereka di masa depan. Dengan jurusan tersebut diharapkan dapat memaksimalkan potensi, bakat atau bakat individu, sehingga dapat memaksimalkan nilai akademik. Sehingga cara yang digunakan untuk menentukan jurusan dengan menerapkan teknik data mining diharapkan dapat membantu siswa dalam menentukan jurusan yang benar sesuai dengan kriteria yang ditetapkan. Teknik data mining digunakan dalam penentuan jurusan ini menggunakan *Algoritma K-Means*.

Penelitian yang di lakukan Eduardus Hardika Sandy Atmaja (2019) dalam *International Journal of Applied Sciences and Smart Technologies Volume 1, Issue 1, pages 33–44* melakukan penelitian dengan judul “*Implementation of K-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta*”. Menyatakan bahwa meningkatnya angka kriminalitas dari hari ke hari perlu menjadi perhatian polisi, sebagai pihak yang bertanggung jawab atas keamanan di masyarakat. Pencegahan kriminalitas usaha harus dilakukan dengan sungguh-sungguh dengan segala ilmu yang dimilikinya. Sehingga cara yang digunakan untuk mengelompokkan data yaitu ke dalam kelompok yang memiliki kemiripan data yang digunakan dalam penelitian ini adalah pola kejahatan yang kemudian dikelompokkan menggunakan algoritma clustering K-medoids. Hasil yang diperoleh dalam hal ini yaitu dilakukan pada tiga kelompok kejahatan yaitu tingkat kejahatan tinggi dengan 4 anggota, tingkat kejahatan sedang dengan 6 anggota dan tingkat kejahatan rendah dengan 8 anggota.

Penelitian yang di lakukan Essau Baker (2020) dalam Jurnal Dinamika Informatika 2020 Universitas PGRI Yogyakarta melakukan penelitian dengan judul “*Implementasi Data Mining Clustering Data Perpustakaan Menggunakan Algoritma K-Means untuk Menentukan Penambahan Koleksi Buku di Perpustakaan UPY*”, menyatakan bahwa permasalahan yang dihadapi yaitu kesulitan dalam menentukan jenis koleksi apa yang seharusnya perlu menjadi prioritas untuk diperbanyak atau tidak, harus menjadi pengetahuan penting bagi para pengelola perpustakaan. Pengelola perpustakaan harus memiliki pengetahuan berdasarkan data pengunjung, sehingga tidak terjadi kesalahan dalam pengadaan. Sehingga cara yang diperlukan adalah sistem yang dapat membantu menentukan sebuah keputusan yang tepat dan cepat. Sehingga akan merancang sebuah sistem implementasi data mining clustering data perpustakaan menggunakan *algoritma K-Means* untuk menentukan penambahan koleksi buku di perpustakaan UPY berbasis website untuk mengatasi masalah yang ada.

Berdasarkan dari enam penelitian relevan diatas dapat disimpulkan bahwa penelitian terdahulu membahas tentang hal yang sejenis dalam proses pengelompokan/*cluster* dengan metode *Algoritma K-Means Clustering*.

Pada penelitian ini digunakan metode *Algoritma K-Means Clustering*, sehingga cocok untuk pengolahan data yang ada pada Dinas Perpustakaan dan Kearsipan Kabupaten Magelang dimana menggunakan Data Transaksi Buku. Perbedaan penelitian ini dengan penelitian relevan yaitu menggunakan jumlah kategori buku berdasarkan kategori buku dan buku yang di pinjam dalam jangka waktu setahun.

2.2 Landasan Teori

2.2.1 Minat Baca

Menurut Ginting (2005) , minat baca adalah bentuk-bentuk perilaku yang terarah guna melakukan kegiatan membaca sebagai tingkat kesenangan yang kuat dalam melakukan kegiatan membaca karena menyenangkan dan memberikan nilai.

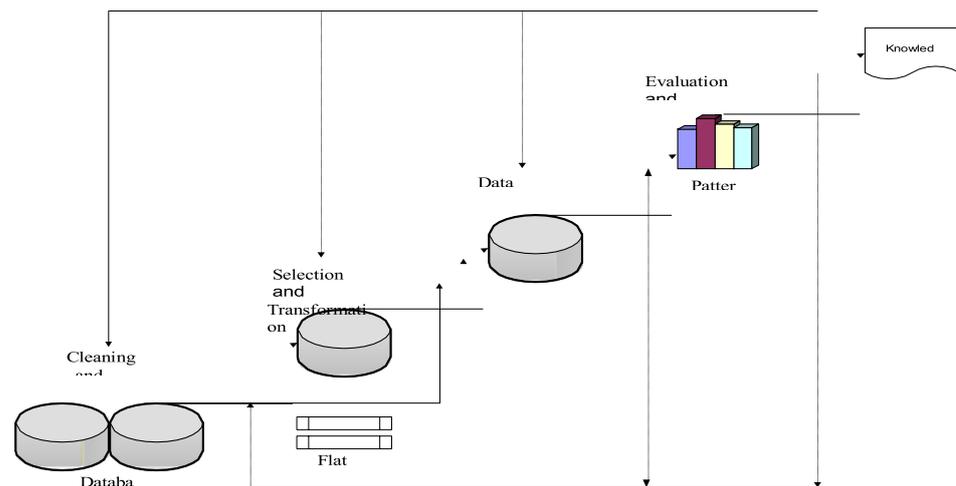
Menurut Widya Safira Azis (2016), Pengelompokan minat baca berdasarkan kriteria buku yang dibaca, buku yang dipinjam, dan juga mempertimbangkan jumlah stok buku yang tersedia dapat membantu dalam proses penambahan koleksi buku yang telah ada pada perpustakaan Utsman Bin Affan UMI. Salah satu cara untuk mengelola data tersebut yaitu menggunakan data mining dengan memanfaatkan metode K- Means. Data buku dikelompokkan menjadi 3 cluster yaitu prioritas, dipertimbangkan, dan tidak diprioritaskan dalam perencanaan penambahan koleksi buku. Hasil dari penelitian ini adalah cluster dengan nilai terbesar pada centroid akhir merupakan cluster yang direkomendasikan dalam perencanaan penambahan koleksi buku.

2.2.1 Data Mining

Data Mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Istilah lain yang sering digunakan diantaranya *knowledge discovery (mining) in databases (KDD)*, *knowledge extraction*, *data atau pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, dan *business intelligence*.

Teknik *data mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Tidak semua pekerjaan pencarian informasi dinyatakan sebagai *Data Mining*.

Data Mining adalah bagian integral dari *knowledge discovery in databases (KDD)*. sebuah langkah dalam proses mencari pola-pola yang terdapat dalam setiap informasi. Langkah-langkah tersebut akan dijelaskan pada gambar 23 (Han, 2011, p6).



Gambar 2. 1 Proses KDD (Sumber : Han 2011)

Proses KDD dalam menghasilkan knowledge dan terdiri dari beberapa tahap:

1. *Data Cleaning*

Untuk menghapus data yang tidak dipakai dan data yang tidak konsisten.

2. *Data Integration*

Berbagai sumber data dapat digabungkan.

3. *Data Selection*

Data yang bersangkutan pada tugas analisis diseleksi dan diambil kembali dari database.

4. *Data Transformation*

Data diubah atau diperkuat menjadi bentuk yang seharusnya untuk diolah dengan menganalisis ringkasan atau jumlah total agregasi.

5. *Data Mining*

Sebuah proses penting di mana metode intelijen diterapkan dengan tujuan untuk mengolah pola-pola data.

6. *Pattern Evaluation*

Untuk mengidentifikasi pola-pola menarik yang menjelaskan mengenai ukuran dasar pengetahuan yang ada.

7. *Knowledge Presentation*

Visualisasi dan teknik representasi knowledge digunakan untuk menyajikan knowledge yang telah diolah untuk pengguna.

2.2.2 *Algoritma K-Means Clustering*

K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.

Terdapat dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu *Hierarchical* dan *Non-Hierarchical*, dan *K-Means* merupakan salah satu metode data *clustering non-hierarchical* atau *Partitional Clustering*.

Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Data clustering menggunakan metode *K-Means Clustering* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

- a. Tentukan jumlah *cluster*
- b. Alokasikan data ke dalam *cluster* secara random

- c. Hitung *centroid*/rata-rata dari data yang ada di masing-masing cluster
- d. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat
- e. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai *threshold* yang ditentukan.

Adapun proses algoritma metode *K-Means Clustering* adalah sebagai berikut:

- a. Tentukan k sebagai jumlah *cluster* yang akan dibentuk, menggunakan metode *elbow criterion* dengan rumus sebagai berikut

$$SSE = \sum_{k=1}^K \sum xi = Sk ||Ni - Ck|| \dots \dots \dots (2.1)$$

SSE = Sum of Square Error

Σ = *Sigma (Total)*

Xi = Jarak data ke-i

Sk = Nilai atribut

Ni = Banyaknya data ke-i

Ck = Candidate k-itemsets (yang potensial untuk menjadi frequent itemsets).

- b. Tentukan k titik pusat *cluster (centroid)* awal yang dilakukan secara random. Penentuan pada *centroid* awal dilakukan secara acak atau random dari objek yang tersedia sebanyak k cluster, untuk menghitung *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n xi}{n} : i = 1,2,3, \dots \dots n \dots \dots \dots (2.2)$$

v = centroid cluster

Σ = *Sigma (Total)*

i = Jarak data ke-i

n = Banyaknya data ke-i

- c. Menghitung jarak dari setiap objek ke masing masing *centroid* dari masing-masing *cluster* menggunakan *Euclidean Distance*, dengan rumus sebagai berikut:

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (X_i - y_i)^2 ; i = 1,2,3, \dots (2.3)}$$

d = *Distance* (jarak)

Σ = *Sigma* (Total)

i = *Jarak data ke-i*

n = *Banyaknya data*

x = *Sumbu x*

y = *Sumbu y*

- d. Alokasikan dari masing-masing objek kedalam *centroid* paling dekat. Untuk pengalokasian objek kedalam masing-masing *cluster* pada saat iterasi secara umum dilakukan dengan cara *hard k-means* dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat *cluster* tersebut.
- e. Lakukan iterasi dan kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan.
- f. Ulangi langkah tiga jika posisi *centroid* baru tidak sama

2.2.3 *Elbow Method*

Ni Putu Eka Merliana et al (2008) mengungkapkan bahwa *Elbow Method* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan cluster terbaik. Dan selain itu persentase perhitungan yang dihasilkan menjadi pembanding antara jumlah cluster yang ditambah.

Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai cluster tersebut yang terbaik. Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (Sum of Square Error) dari masing-masing nilai cluster. Karena semakin besar jumlah cluster K maka nilai SSE akan semakin kecil

2.2.4 Jupyter

Jupyter merupakan perangkat lunak yang bersifat open source dan servis dalam komputasi yang interaktif dalam berbagai macam bahasa pemrograman. *JupyterLab* adalah interactive development environment yang berbasis web untuk *Jupyter Notebooks* kode program, dan data. *JupyterLab* fleksibel dalam hal mendukung workflow untuk data sains, komputasi ilmiah, dan *machine learning*. *JupyterLab* juga bersifat ekstensibel dan modular. Jupyter notebook adalah sebuah aplikasi web yang bersifat open source yang diperuntukan dalam data cleaning dan transformasinya, simulasi angka, visualisasi data, pemodelan statistic, *machine learning* dll.

2.2.6 Flask

Flask adalah sebuah aplikasi dalam bahasa pemrograman *python* yang menggunakan *toolkit Werkzeug* dan *template Jinja2*. Dalam sejarah perkembangannya, *flask* pertama kali diliris pada April 2010 dibuat oleh Armin Ronacher. Kelebihan dari *flask* adalah *framework website* yang lebih mudah dipahami dan sederhana dibandingkan *framework* lainnya. *Flask* menyederhanakan inti *framework*-nya sehingga menjadi lebih ringan dan cepat. Oleh karena itu *flask* disebut dengan *microframework*. *Flask* juga memiliki fungsi yang bisa dikembangkan sesuai kebutuhan (Samudera, 2015).

Flask terbagi menjadi dua komponen utama yaitu *Werkzeug* dan *Jinja2*. *Werkzeug* berperan dalam menyediakan *routing*, *debugging*, dan *Web Server Interface Gateway (WSGI)*, sedangkan *Jinja2* sebagai *template engine*. Tidak seperti *Django*, *flask* sendiri tidak mendukung fungsi-fungsi seperti *database access*, *user authentication*, dan perangkat “*high-level*” lainnya. Namun *flask* dapat dikombinasikan dengan berbagai *extention* untuk menambah berbagai fungsi (Relan, 2019). Tahapan singkat mengenai pembuatan web sederhana menggunakan *flask*, dengan asumsi sudah meng-install aplikasi python dan anaconda terlebih dahulu:

1. Setting *Enviroment* atau yang sering disebut dengan mengatur virtual *environment* secara sederhana yaitu mengatur peletakan aplikasi *python* agar dapat dibaca oleh program *flask* terutama saat melakukan tahap *hosting website*.

2. Install *flask* melalui *pip* pada *command prompt python*. Pada *python*, *package* seperti *flask* tersedia pada *public respiratory* artinya semua orang dapat *download* dan *meng-installnya*. *Official python package respiratory* pada *python* dinamakan dengan *PyPI (Python Package Index)*. Untuk *meng-install package* dari *PyPI* cukup mudah, dapat menggunakan alat “*pip*” dengan menuliskan code “*pip install* ”. Jadi cara *meng-install flask* dapat menggunakan cara “*pip install flask*” (Grinberg, 2017).

3. Membuat code server

```
1  from flask import Flask
2  application = Flask(__name__)
3
4  @application.route('/')
5  def index():
6      return '<h1>Hello World</h1>'
7
8  if __name__ == '__main__':
9      application.run(debug=True)
```

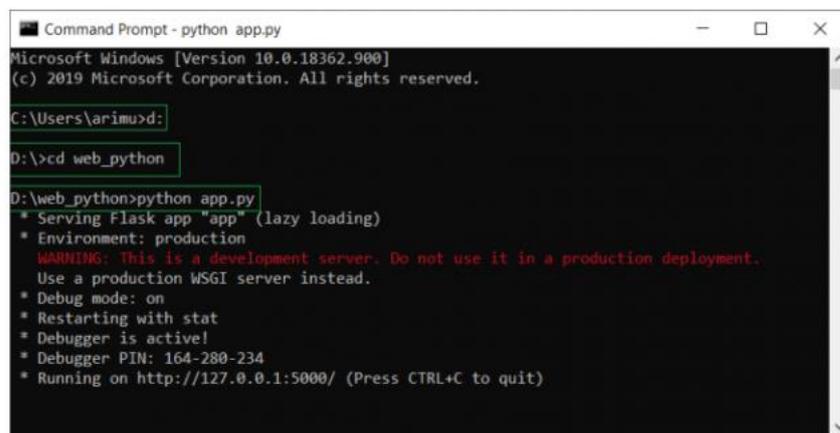
Gambar 2. 2 Server Code Flask (Sumber : Villalon 2018)

Code atau *script* penting *flask* terbagi dalam contoh pada gambar 2.1. Baris pertama melakukan *import module flask* untuk membuat *flask web server*

dari *module flask*. Baris ketiga disini mempresentasikan web aplikasi, jadi membuat flask agar memahami bahwa program ini adalah web aplikasi. Baris kelima merepresentasikan halaman web utama. Kemudian baris keenam dan ketujuh merupakan ketika user memasuki website dengan link atau URLs “/” maka akan masuk ke halaman dengan tulisan “*Hello World!*”. Baris kesembilan arti “*main*” sendiri yaitu utama artinya jika kita menjalankan script tersebut maka python akan mengutakan script tersebut untuk berjalan dan mencegah script lain untuk berjalan. Yang terakhir baris kesepuluh code ini akan menjalankan aplikasi, dan “*debug = True*” artinya jika terjadi error maka python akan menampilkannya pada halaman web untuk memudahkan dalam memantau dan menangani error (Villalon, 2018).

4. Membuka URL hasil *coding*

Setelah membuat *code server* dan disimpan dengan nama contoh “*main.py*”. Maka selanjutnya membuka *script* tersebut pada *command prompt* dengan cara mengetik “*py main.py*” atau “*python main.py*”. sehingga muncul output berikut pada gambar 2.2. Dan jalankan <http://127.0.0.1:5000> melalui aplikasi browser.



```
Command Prompt - python app.py
Microsoft Windows [Version 10.0.18362.900]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\arimu>d:
D:\>cd web_python
D:\web_python>python app.py
* Serving Flask app "app" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
* Debugger is active!
* Debugger PIN: 164-280-234
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Gambar 2. 3 Output Flask Pada Command Promt

Secara *front-end* digunakan bahasa markup HTML maupun CSS sebagai tampilan, sedangkan secara *back-end* menggunakan framework *flask*.

2.2.7 Silhouette Coefficient

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan cluster, seberapa baik suatu objek ditempatkan dalam suatu cluster. Metode ini merupakan gabungan dari metode cohesion dan separation. Tahapan perhitungan Silhouette Coefficient adalah sebagai berikut:

1. Hitung rata-rata jarak dari suatu dokumen misalkan i dengan semua dokumen lain yang berada dalam satu cluster

$$a(i) = \frac{1}{|A-1|} \sum_{j \in A, j \neq i} d(i,j) \dots\dots\dots(2.4)$$

dengan j adalah dokumen lain dalam satu cluster A dan $d(i,j)$ adalah jarak antara dokumen i dengan j .

2. Hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di cluster lain, dan diambil nilai terkecilnya

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i,j) \dots\dots\dots(2.5)$$

dengan $d(i,C)$ adalah jarak rata-rata dokumen i dengan semua objek pada cluster lain C dimana $A \neq C$.

$$b(i) = \min_{C \neq A} d(i, C) \dots\dots\dots(2.6)$$

3. Nilai Silhouette Coefficient nya adalah :

$$S(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \dots\dots\dots(2.7)$$

Menurut L.Kaufmann dan P.J. Rousseeuw (1990) Nilai silhouette coefficient dapat bervariasi antara -1 hingga 1. Hasil clustering dikatakan baik jika nilai silhouette coefficient bernilai positif ($a(i) < b(i)$) dan $a(i)$ mendekati 0, sehingga akan menghasilkan nilai silhouette coefficient yang maksimum yaitu 1 saat $a(i) = 0$. Dengan demikian, jika $s(i) = 1$ berarti objek i sudah berada dalam cluster yang tepat. Jika nilai $s(i) = 0$ maka objek i berada diantara dua cluster sehingga objek tersebut tidak jelas harus dimasukkan ke dalam cluster A atau cluster B . Akan tetapi, jika $s(i) = -1$ artinya struktur cluster yang dihasilkan overlapping, sehingga objek i lebih tepat dimasukkan ke dalam cluster yang lain

Nilai rata-rata silhouette dari tiap objek dalam suatu cluster adalah suatu ukuran yang menunjukkan seberapa ketat/fit dan optimal sekelompok data berada pada cluster tersebut. Ukuran nilai *silhouette coefficient* bisa dilihat pada tabel 2.1

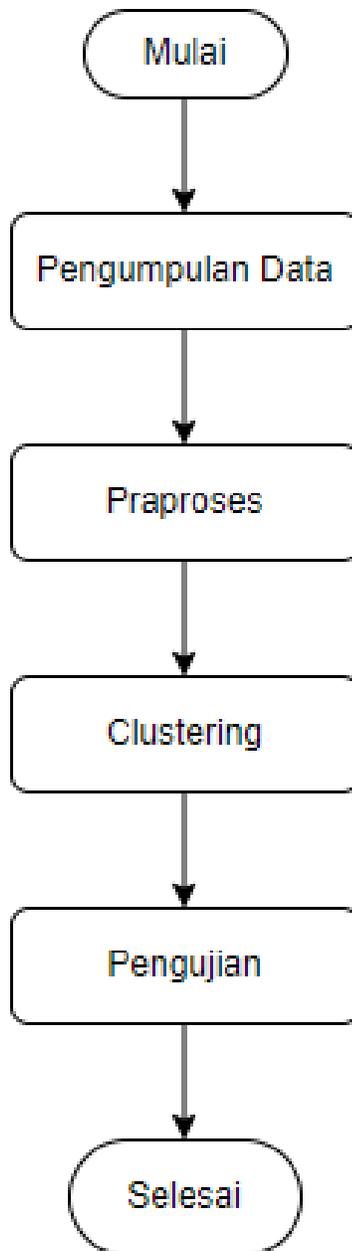
Tabel 2. 1 Ukuran Nilai Silhouette Coefficient L.Kaufmann & P.J. Rousseeuw (1990)

Nilai Silhouette Coefficient	Kekuatan Cluster
$0.7 < SC \leq 1$	struktur yang kuat
$0.5 < SC \leq 0.7$	struktur sedang
$0.25 < SC \leq 0.5$	struktur lemah
$SC \leq 0.25$	tidak ada struktur

BAB 3 METODOLOGI PENELITIAN

3.1 Prosedur Penelitian

Prosedur penelitian yang digunakan menggunakan metode *K-Means Clustering*. Prosedur penelitian yang dilakukan terlihat seperti pada Gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

Adapun prosedur penelitian yang dilakukan seperti pada gambar 3.1 ini adalah sebagai berikut :

1. Pengumpulan Data

Pada tahap ini adalah meminta data , data yang di dapatkan adalah data transaksi peminjaman dan pengembalian buku dari Januari 2020 sampai Desember 2020 dari Aplikasi *Inlislite* di di Dinas Perpustakaan dan Kearsipan Kabupaten Magelang kumpulkan terlebih dahulu.

2. Praproses

Setelah data didapat, kemudian di lakukan tahap praproses. Pada tahapan praproses ini yaitu :

- a. Tahap *Pre-Processing*. Proses *Pre-Processing* mencakup Data Outlier. Dimana data yang menyimpang terlalu jauh dari data yang lainnya dalam suatu rangkaian data (pencilan).
- b. Setelah tahapan *Data Outlier* dilakukan Scaled. Agar data dapat kembali seperti jenis data diawal sebelum di array kan. *Min-Max Scaler* bekerja dengan *scaling* data/menyesuaikan data dalam rentang/*range* tertentu (*range* nilai *minimum* hingga nilai *maksimum*), dengan rentang yang biasa digunakan adalah 0 hingga 1.

3. *Clustering*

Pada tahap ini mengelompokkan data yang ada ke dalam beberapa kelompok. Dalam tahapan ini akan dibuat melakukan *Clustering* Minat Baca. Untuk langkah-langkah melakukan *Clustering* dengan metode *Algoritma* K-Means adalah sebagai berikut:

4. *Pengujian*

Pada tahap ini merupakan proses memberikan pengujian atas hasil *clustering*. Dimana pada penelitian ini menggunakan *Silhouette Coefficient*

3.2 Metode Pengumpulan Data

Pada Metode Pengumpulan Data, Data yang digunakan sebagai objek penelitian dalam penelitian ini merupakan data sekunder yang diperoleh dari Aplikasi *Inlislite* Dinas Perpustakaan dan Kearsipan Kabupaten Magelang. Data yang diambil adalah data transaksi buku dari Januari 2020 – Desember 2020.

3.3 Metode Pengolahan Data

Adapun pengolahan data mining yang dilakukan pada penelitian ini menggunakan jupyter notebook. Dengan tahapan sebagai berikut :

1. Menentukan atribut yang akan di cluster kan

Pada tahap ini bertujuan untuk menentukan atribut sebelum di clusterkan yaitu untuk melihat atribut mana saja yang masuk tahap clustering.

2. Memvisualkan persebaran data

Pada tahap ini bertujuan untuk untuk melihat persebaran data dari data yang divisualkan.

3. Mengubah Variabel Data Frame Menjadi Array

Tujuan dari menstandarkan kembali ukuran variabel yaitu agar data dapat kembali seperti jenis data diawal sebelum di array kan

4. Menentukan Nilai K

Pada tahap ini bertujuan untuk mencari nilai K atau banyak nya n cluster didapatkan dari *elbow method* atau metode siku.

5. Menstandarkan ukuran variabel

Pada tahap ini bertujuan untuk menstandarkan kembali ukuran variabel yaitu agar ukuran variabel menyesuaikan data dalam rentang/*range* tertentu (*range* nilai *minimum* hingga nilai *maksimum*), dengan rentang yang biasa digunakan adalah 0 hingga 1.

6. Menentukan Kmeans

Pada tahap ini bertujuan untuk memasukkan banyaknya nilai k pada kmeans. Yang didapat dari menentukan nilai K menggunakan metode elbow sehingga bertujuan mencari nilai cluster terbaik.

7. Menampilkan hasil cluster

Pada tahap ini bertujuan untuk menampilkan hasil cluster setelah menentukan banyaknya nilai k pada kmeans

8. Menambahkan kolom cluster

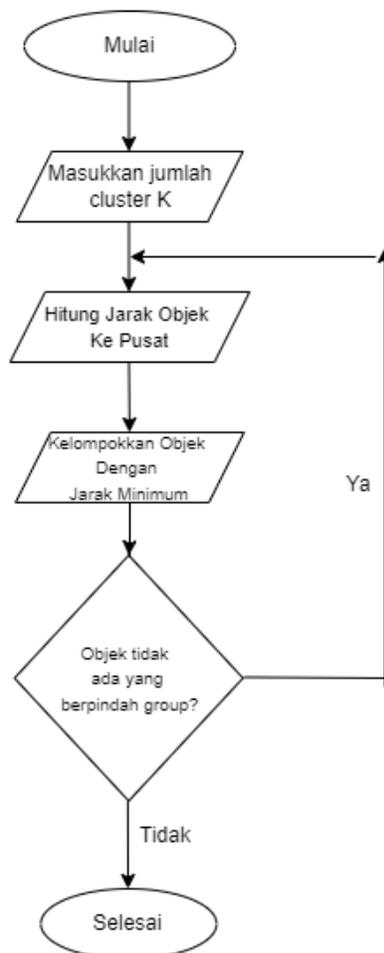
Pada tahap ini bertujuan untuk menambahkan kolom yang bernama

kolom cluster setelah diketahui hasil cluster dari tahap sebelumnya.

9. Visualisasikan hasil cluster

Pada tahap ini bertujuan untuk mengetahui persebaran hasil cluster setelah dilakukan visualisasi menggunakan diagram lingkaran atau kartesian dengan nilai x dan y.

Adapun tahapan pengerjaan metode K-Means Clustering dalam penelitian ini seperti dalam flowchart pada gambar 3.2



Gambar 3. 2 Algoritma K-Means Clustering

3.4 Prosedur Pengujian Metode

Prosedur pengujian metode menggunakan metode *Silhouette Coefficient*, *Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan cluster, seberapa baik suatu objek ditempatkan dalam suatu cluster.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan yaitu :

- a. Minat baca masyarakat di Kabupaten Magelang berdasarkan pengelompokan menggunakan *K-Means Clustering* tergolong rendah. Hal ini terjadi baik dikelompokkan menjadi 2 atau 3 kelompok.
- b. Berdasarkan pengujian menggunakan *Silhouette Coefficient, Clustering* dengan 2 kelompok lebih kuat dari *Clustering* menggunakan 3 kelompok.
- c. Berdasarkan pengelompokan kategori buku. Ketertarikan minat masyarakat yaitu kategori buku kesusasteraan dengan 1552 kali pinjaman.

5.2 Saran

Saran yang diperlukan untuk kedepannya adalah sebagai berikut :

- a. Penelitian selanjutnya dapat dikembangkan pada jenis data yang serupa dengan metode yang lebih baik seperti menggunakan algoritma *Fuzzy C-Means*.
- b. Hasil *clustering* yang terbentuk dapat dikembangkan menjadi pengelompokan yang berguna untuk pengunjung di Dinas tersebut