



**PENENTUAN *CLUSTER* HIERARKI OPTIMAL PADA
BERBAGAI *LINKAGE* DAN JARAK (STUDI PERILAKU
KREDITUR BANK X)**

SKRIPSI

Oleh:

FAUZAN RISMANDA ROSANTA PUTRA

155090501111030



PROGRAM STUDI SARJANA STATISTIKA

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS BRAWIJAYA

MALANG

2021

LEMBAR PENGESAHAN SKRIPSI

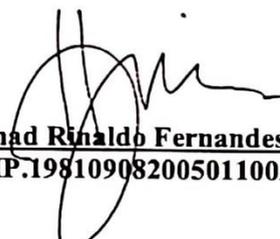
**PENENTUAN CLUSTER HIERARKI OPTIMAL PADA
BERBAGAI LINKAGE DAN JARAK (STUDI PERILAKU
KREDITUR BANK X)**

Oleh:

**FAUZAN RISMANDA ROSANTA PUTRA
155090501111030**

Setelah dipertahankan di depan Majelis Penguji
Pada 22 Desember 2021
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Statistika dalam bidang Statistika

Dosen Pembimbing,


Dr. Adji Achmad Rinaldo Fernandes, S.Si., M.Sc.
NIP.198109082005011002

Mengetahui,
Ketua Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Brawijaya,




Rahma Fitriani, S.Si., M.Sc., Ph.D.
NIP. 197603281999032001

LEMBAR PERTANYAAN

Saya yang bertanda tangana di bawah ini:

NAMA : Fauzan Rismanda Rosanta Putra

NIM : 155090501111030

PROGRAM STUDI : STATISTIKA

SKRIPSI BERJUDUL :

PENENTUAN *CLUSTER* HIERARKI OPTIMAL PADA BERBAGAI *LINKAGE* DAN JARAK (STUDI PERILAKU KREDITUR BANK X)

Dengan ini menyatakan bahwa:

1. Isi dari skripsi yang saya buat adalah benar-benar karya saya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termasuk di isi dan tertulis di daftar pustaka dalam Skripsi ini.
2. Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung risiko yang akan saya terima.

Demikian perppnyataan ini dibuat dengan segala kesadaran.

Malang, 22 Desember 2021

Yang menyatakan,

Fauzan Rismanda Rosanta Putra

NIM. 155090501111030



PENENTUAN *CLUSTER* HIERARKI OPTIMAL PADA BEBRBAGAI *LINKAGE* DAN JARAK (STUDI PERILAKU KREDITUR BANK X)

ABSTRAK

Analisis *cluster* merupakan teknik multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik. Tujuan dari penelitian ini adalah melakukan pengelompokan nasabah kreditur KPR di Bank X berdasarkan penilaian 6 K. dengan menggunakan metode pengukuran jarak *Euclidean distance*, *Manhattan distance*, dan *Mahalanobis distance*. Sedangkan untuk metode pengelompokan menggunakan *Ward's method*, *Complete Linkage*, dan *Average Linkage*. Beberapa metode pengukuran jarak membutuhkan syarat tertentu, seperti *Euclidean distance* dan *Manhattan distance* yang mengharuskan tidak terdapat korelasi antar variabel yang digunakan. Pada penelitian ini akan membandingkan hasil pengelompokan menggunakan ketiga metode berdasarkan rasio simpangan baku antar *cluster* dan dalam *cluster*. Teknik pengambilan sampel menggunakan metode *nonprobability sampling* dengan basis *purposive sampling*. Ukuran sampel pada penelitian ini adalah 300 nasabah. Rasio simpangan baku antar *cluster* dan dalam *cluster* menunjukkan bahwa hasil pengelompokan metode *Ward* dengan *Mahalanobis distance* lebih baik dibandingkan hasil pengelompokan menggunakan metode lainnya. Hasil analisis *cluster* menggunakan metode *Ward* dengan *Mahalanobis distance* adalah *cluster* satu terdiri dari 137 nasabah dengan kepatuhan sebagai aspek tertinggi, *cluster* dua terdiri dari 119 nasabah dengan *condition* sebagai aspek tertinggi, sedangkan *cluster* tiga terdiri dari 44 nasabah dengan *capital* sebagai aspek tertinggi.

Kata Kunci : Analisis *Cluster*, *Euclidean distance*, *Manhattan distance*, *Mahalanobis distance*, *Ward's Method*, *Complete Linkage Method*, *Average Linkage Method*



DETERMINATION OF THE OPTIMAL CLUSTER HIERARCHY WITH VARIOUS LINKAGE AND DISTANCE APPROACHES (STUDY OF CREDITORS BANK X)

ABSTRACT

Cluster analysis is a multivariate technique that aims to classify objects based on characteristic similarities. The purpose of this study is to classify creditors in Bank X based on evaluation of 6 K using the Euclidean distance, Manhattan distance, and Mahalanobis distance as distance measurements method. Then Ward's method, Complete Linkage and Average Linkage method as linkage methods. Some measurements methods needs specific condition, like Single Linkage and Manhattan Linkage that needs the data free from correlation between variables. This research will compare the results of grouping of all methods based on standard deviation ratio between groups. The sampling technique used in this study is the nonprobability sampling method on the basis of purposive sampling. The sample size in this study was 300 tourists. The ratio of standard deviation between clusters and in clusters, the results of Ward method with Mahalanobis distance is better than the other methods. Ward method with Mahalanobis distance group results are cluster one, consisting of 137 clients with the highest aspect is obedient, cluster two consists of 119 clients with the highest aspect is condition, and cluster three consists of 44 clients with the highest aspect is capital.

Keywords : Cluster Analysis, Euclidean distance, Manhattan distance, Mahalanobis distance, Ward's Method, Complete Linkage, Average Linkage.



DAFTAR ISI

	Hal
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	iii
LEMBAR PERTANYAAN.....	v
ABSTRAK.....	vii
ABSTRACT.....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL.....	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	4
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1. Analisis <i>Cluster</i>	7
2.2. Metode Pengukuran Jarak.....	8
2.2.1. <i>Euclidean Distance</i>	9
2.2.2. <i>City-Block (Manhattan) Distance</i>	9
2.2.3. <i>Mahalanobis Distance</i>	10
2.3. Metode Pengelompokan.....	10
2.3.1. Metode <i>Ward</i>	11
2.3.2. <i>Complete Linkage Method</i>	11
2.3.3. <i>Average Linkage Method</i>	12
2.4. Pemeriksaan Multikolinearitas.....	13
2.4.1. <i>Variance Inflation Factor (VIF)</i>	13
2.4.2. Analisis Komponen Utama (AKU).....	13
2.5. Indeks Validitas <i>Cluster</i>	17
2.5.1. Indeks <i>Dunn (D)</i>	17
2.5.2. Indeks <i>Davies-Bouldin (DB)</i>	17
2.5.3. Indeks <i>Global Silhouette (Gsu)</i>	18
2.5.4. Indeks Gabungan.....	19



2.6.	Penentuan Kebaikan Metode	19
2.7.	Variabel dan Data Penelitian	20
2.8.	Metode Pengukuran Variabel	22
2.9.	Pemeriksaan Instrumen Penelitian	22
2.9.1.	Pemeriksaan Validitas	22
2.9.2.	Pemeriksaan Reliabilitas	24
2.10.	Variabel Penelitian	24
2.10.1.	Karakter	24
2.10.2.	Kapasitas	25
2.10.3.	Kapital	25
2.10.4.	Kekayaan	25
2.10.5.	Kondisi	25
2.10.6.	Kepatuhan	26
BAB III METODE PENELITIAN		27
3.1.	Sumber Data	27
3.2.	Lokasi dan Waktu Penelitian	27
3.3.	Populasi dan Sampel	27
3.4.	Variabel Penelitian	28
3.5.	Kisi-Kisi Instrumentasi	28
3.6.	Evaluasi Instrumen Penelitian	32
3.7.	Metode Penelitian	35
3.8.	Diagram Alir Penelitian	37
BAB IV HASIL DAN PEMBAHASAN		39
4.1.	Uji Multikolinearitas	39
4.2.	Hasil Analisis <i>Cluster</i>	39
4.2.1.	Hasil Analisis <i>Cluster</i> menggunakan <i>Euclidean distance</i> dengan metode <i>Ward</i>	40
4.2.2.	Hasil Analisis <i>Cluster</i> Menggunakan <i>Euclidean Distance</i> dengan <i>Complete Linkage</i>	42
4.2.3.	Hasil Analisis <i>Cluster</i> Menggunakan <i>Euclidean Distance</i> dengan <i>Average Linkage</i>	44
4.2.4.	Hasil Analisis <i>Cluster</i> Menggunakan <i>Manhattan</i> <i>Distance</i> dengan metode <i>Ward</i>	46
4.2.5.	Hasil Analisis <i>Cluster</i> Menggunakan <i>Manhattan</i> <i>Distance</i> dengan <i>Complete Linkage</i>	48



DAFTAR GAMBAR

	Hal
Gambar 2. 1. Dendogram cluster hierarki	7
Gambar 2. 2. <i>Complete Linkage Method</i>	12
Gambar 2. 3. <i>Average Linkage Method</i>	12
Gambar 3.1. Diagram Alir Penelitian	37
Gambar 4. 1. Dendogram <i>Euclidean distance</i> dengan metode <i>Ward</i>	42
Gambar 4. 2. Dendogram <i>Euclidean Distance</i> dengan <i>Complete Linkage</i>	44
Gambar 4. 3. Dendogram <i>Euclidean distance</i> dengan <i>Average Linkage</i>	46
Gambar 4. 4. Dendogram <i>Manhattan distance</i> dengan metode <i>Ward</i>	48
Gambar 4. 5. Dendogram <i>Manhattan Distance</i> dengan <i>Complete Linkage</i>	50
Gambar 4. 6. Dendogram <i>Manhattan Distance</i> dengan <i>Average Linkage</i>	52
Gambar 4. 7. Dendogram <i>Mahalanobis Distance</i> dengan metode <i>Ward</i>	54
Gambar 4. 8. Rata-Rata Variabel Kepuasan Wisatawan dari Hasil <i>Cluster</i> menggunakan <i>Mahalanobis Linkage</i> dengan metode <i>Ward</i>	61



DAFTAR TABEL

	Hal
Tabel 3. 1. Struktur Data Penelitian.....	27
Tabel 3. 2. Kisi-Kisi Instrumen Penelitian	28
Tabel 3. 3. Uji Validitas Instrumen Penelitian.....	32
Tabel 3. 4. Pemeriksaan Reliabilitas Kuesioner	34
Tabel 4. 1. Hasil VIF dengan Y_1	39
Tabel 4. 2. Indeks <i>Dunn, Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Euclidean Distance</i> dengan metode <i>Ward</i>	41
Tabel 4. 3. Hasil Peringkat Indeks Gabungan Jarak <i>Euclidean</i> dengan metode <i>Ward</i>	41
Tabel 4. 4. Anggota Tiga <i>Cluster</i> Menggunakan <i>Euclidean distance</i> dengan metode <i>Ward</i>	42
Tabel 4. 5. Indeks <i>Dunn, Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Euclidean Distance</i> dengan <i>Complete Linkage</i>	43
Tabel 4. 6. Hasil Peringkat Indeks Gabungan Jarak <i>Euclidean</i> dengan <i>Complete Linkage</i>	43
Tabel 4. 7. Jumlah Anggota dari Empat <i>Cluster</i> Menggunakan <i>Euclidean Distance</i> dengan <i>Complete Linkage</i>	44
Tabel 4. 8. Indeks <i>Dunn, Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Euclidean Distance</i> dengan <i>Average Linkage</i>	45
Tabel 4. 9. Hasil Peringkat Indeks Gabungan Jarak <i>Euclidean</i> dengan <i>Average Linkage</i>	45
Tabel 4. 10. Jumlah Anggota dari Tiga <i>cluster</i> menggunakan <i>Euclidean distance</i> dengan <i>Average Linkage</i>	46
Tabel 4. 11. Indeks <i>Dunn, Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Manhattan Distance</i> dengan Metode <i>Ward</i>	47
Tabel 4. 12. Hasil Peringkat Indeks Gabungan Jarak <i>Manhattan</i> dengan Metode <i>Ward</i>	47
Tabel 4. 13. Jumlah Anggota dari Tiga <i>cluster</i> menggunakan <i>Manhattan distance</i> dengan Metode <i>Ward</i>	48
Tabel 4. 14. Indeks <i>Dunn, Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Manhattan Distance</i> dengan <i>Complete Linkage</i>	49
Tabel 4. 15. Hasil Peringkat Indeks Gabungan Jarak <i>Manhattan</i> dengan <i>Complete Linkage</i>	49



Tabel 4. 16. Jumlah Anggota pada 4 <i>Cluster</i> dengan <i>Manhattan Distance</i> dan <i>Complete Linkage</i>	50
Tabel 4. 17. Indeks <i>Dunn</i> , <i>Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Manhattan Distance</i> dengan <i>Average Linkage</i>	51
Tabel 4. 18. Hasil Peringkat Indeks Gabungan Jarak <i>Manhattan</i> dengan <i>Average Linkage</i>	51
Tabel 4. 19. Jumlah Anggota pada 3 <i>Cluster</i> dengan <i>Manhattan Distance</i> dan <i>Average Linkage</i>	52
Tabel 4. 20. Indeks <i>Dunn</i> , <i>Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Mahalanobis Distance</i> dengan metode <i>Ward</i>	53
Tabel 4. 21. Hasil Peringkat Indeks Gabungan Jarak <i>Mahalanobis</i> dengan Metode <i>Ward</i>	53
Tabel 4. 22. Jumlah Anggota pada 3 <i>Cluster</i> dengan <i>Mahalanobis Distance</i> dan Metode <i>Ward</i>	54
Tabel 4. 23. Indeks <i>Dunn</i> , <i>Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Mahalanobis Distance</i> dengan <i>Complete Linkage</i>	55
Tabel 4. 24. Hasil Peringkat Indeks Gabungan Jarak <i>Mahalanobis</i> dengan <i>Complete Linkage</i>	55
Tabel 4. 25. Jumlah Anggota dari Tiga <i>Cluster</i> Menggunakan <i>Mahalanobis Distance</i> dengan <i>Complete Linkage</i>	56
Tabel 4. 26. Indeks <i>Dunn</i> , <i>Davies Bouldin</i> , dan <i>Silhouette</i> pada <i>Mahalanobis Distance</i> dengan <i>Average Linkage</i>	56
Tabel 4. 27. Hasil Peringkat Indeks Gabungan Jarak <i>Mahalanobis</i> dengan <i>Average Linkage</i>	57
Tabel 4. 28. Jumlah Anggota dari Tiga <i>Cluster</i> Menggunakan <i>Mahalanobis Distance</i> dengan <i>Average Linkage</i>	58
Tabel 4. 29. Rasio Simpangan Baku Dalam <i>Cluster</i> dan Antar <i>Cluster</i>	59
Tabel 4. 30. Rata-Rata Variabel pada Setiap <i>Cluster</i>	60



BAB I PENDAHULUAN

1.1. Latar Belakang

Rumah atau tempat tinggal adalah salah satu kebutuhan primer bagi setiap orang. Namun, pada kenyataannya tidak semua orang dapat membeli rumah yang layak secara tunai. Hal ini disebabkan oleh harga rumah yang terus naik dikarenakan oleh permintaan pasar terhadap rumah semakin bertambah sedangkan ketersediaan lahan yang sangat terbatas. Pendapatan yang cenderung tetap (stagnan) semakin mempersulit masyarakat untuk memiliki rumah, terutama bagi golongan menengah kebawah. Sehingga diperlukan sebuah solusi untuk masalah ini. Dalam hal ini perbankan memiliki peran yang sangat penting.

Undang-Undang Nomor 10 Tahun 1998 tentang perbankan menjelaskan bahwa bank sebagai badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup masyarakat. Salah satu fungsi bank adalah sebagai lembaga perantara (intermediasi) terutama dalam hal kegiatan perkreditan yang dapat membantu pertumbuhan ekonomi secara keseluruhan bagi masyarakat. Guna mendorong pertumbuhan ekonomi, sektor perbankan memiliki peran yang sangat penting dalam membantu masyarakat dalam hal keuangan termasuk memberikan fasilitas kredit.

Salah satu fasilitas kredit yang ditawarkan oleh perbankan kepada nasabahnya adalah Kredit Pemilikan Rumah (KPR). Kredit Pemilikan Rumah adalah sistem pembelian rumah dengan skema pembayaran sebesar persentase tertentu dari harga rumah. Artinya nasabah tidak perlu menyediakan uang sejumlah harga rumah yang akan dibeli, namun hanya perlu membayar uang muka dan selanjutnya membayar angsuran setiap bulan sesuai dengan nominal yang sudah disepakati. Namun pada praktiknya, KPR memiliki tingkat risiko yang tinggi bagi bank, risiko ini dapat disebabkan oleh pihak internal (bank) atau pihak eksternal (nasabah). Upaya dalam mengurangi persentase terjadinya risiko yang disebabkan oleh nasabah, bank dapat



melakukan identifikasi dan analisis statistika terhadap nasabah dengan penerapan prinsip perkreditan 6 K, yaitu karakter, kapital (modal), kapasitas, kekayaan, kondisi, dan kepatuhan nasabah dalam membayar angsuran. Menurut Walpole, dkk. (2012), statistika dapat digunakan untuk membantu dalam proses pembuatan sebuah keputusan ilmiah terkait ketidakpastian dan keragaman yang dihadapi. Statistika adalah bidang ilmu pengetahuan yang berhubungan dengan cara-cara pengumpulan, pengolahan dan pengambilan keputusan yang didasarkan pada fakta dan pengamatan yang dilakukan (Sudjana, 2004).

Salah satu analisis dalam statistika yang termasuk dalam analisis multivariat adalah analisis *cluster* yang memiliki tujuan utama untuk mengelompokkan objek yang diteliti menjadi beberapa *cluster* (kelompok) berdasarkan kesamaan karakteristik dari objek tersebut. Analisis ini akan mengelompokkan objek yang relatif sama sehingga sebuah *cluster* memiliki objek-objek yang homogen dengan tetap menjaga keheterogenan antar *cluster* (Hair, dkk. 2010).

Pada analisis *cluster*, dibutuhkan sebuah metode pengukuran kesamaan untuk mengukur tingkat keheterogenan/kesamaan antar objek, sehingga dapat digunakan sebagai dasar pembentukan *cluster* baru. Metode pengukuran kesamaan yang paling sering digunakan adalah metode pengukuran jarak. Terdapat beberapa metode pengukuran jarak, diantaranya *Euclidean distance*, *City Block (Manhattan) distance*, dan *Mahalanobis distance*. Pembentukan *cluster* baru dilakukan dengan menggunakan metode pengelompokan (*linkage*) berdasarkan ukuran jarak antar objek. Terdapat beberapa metode pengelompokan, seperti *Ward method*, *Complete Linkage method*, *Average Linkage method*.

Setiap metode pengukuran jarak memiliki kekurangan dan kelebihan. Kelebihan *Euclidean distance* adalah relatif lebih mudah dilakukan penghitungannya tetapi sangat sensitif terhadap besarnya sampel dan korelasi antar variabel. *Manhattan distance* merupakan metode pengukuran jarak yang paling mudah untuk diterapkan, namun *cluster* yang terbentuk akan tidak valid jika data memiliki korelasi. Pada *Mahalanobis distance*, korelasi antar variabel yang mungkin ada



tidak diperhitungkan dan setiap variabel diberikan bobot yang sama, tetapi jika data tidak memiliki korelasi maka *Mahalanobis distance* akan dianggap sama dengan *Euclidean distance* yang distandarisasi.

Metode pengelompokan juga memiliki kelebihan dan kekurangan. Kelebihan dari *Complete Linkage* adalah tidak terpengaruh oleh adanya *outliers*, tetapi cenderung membuat *cluster* menjadi berukuran besar. Sedangkan *Average Linkage* akan menggunakan seluruh nilai jarak yang ada di dalam *cluster* sehingga membuat *cluster* yang terbentuk menjadi lebih baik, kekurangannya adalah jika data yang digunakan tidak memiliki korelasi antar individu maka *Mahalanobis distance* akan dianggap sama dengan *Euclidean distance* yang distandarisasi.

Penelitian kedua dilakukan oleh Putri (2017). Metode pengelompokan yang digunakan pada penelitian tersebut adalah *Single Linkage method*, *Complete Linkage method* dan *Average Linkage method* dengan *Euclidean distance* sebagai metode pengukuran jarak. Kesimpulan yang didapat yaitu *Complete Linkage Method* merupakan metode pengelompokan terbaik.

Penelitian mengenai analisis *cluster* telah dilakukan oleh Cahyoningtyas (2019). Pada penelitian tersebut menggunakan *Euclidean distance* dengan *Average Linkage method* dan *Ward method*. Kesimpulannya yaitu *Ward method* merupakan metode pengelompokan terbaik.

Penelitian mengenai analisis *cluster* juga dilakukan oleh Badung (2021). Penelitian tersebut dilakukan untuk menemukan ukuran jarak yang terbaik antara *Euclidean distance*, *Manhattan distance* dan *Mahalanobis distance*. Kesimpulannya yaitu *Mahalanobis distance* merupakan metode pengukuran jarak terbaik.

Merujuk penelitian sebelumnya, penelitian ini bertujuan untuk menemukan kombinasi paling optimal antara tiga metode pengukuran jarak, yaitu *Euclidean distance*, *Manhattan distance* dan *Mahalanobis distance* dengan tiga metode pengelompokan, yaitu *Complete Linkage method*, *Average Linkage method*, dan *Ward method* pada nasabah KPR Bank X. Data yang digunakan adalah data primer yaitu data penilaian variabel 6 K pada Bank X tahun 2021.



1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah:

- 1) Bagaimana pengelompokan nasabah KPR Bank X berdasarkan penilaian variabel 6 K menggunakan kombinasi antara metode pengukuran jarak *Euclidean distance*, *City-Block (Manhattan) distance*, dan *Mahalanobis distance* dengan metode pengelompokan *Ward's method*, *Complete Linkage method*, dan *Average Linkage method*?
- 2) Bagaimana perbandingan *Euclidean distance*, *City-Block (Manhattan) distance* dan *Mahalanobis distance* dengan metode pengelompokan *Ward's method*, *Complete Linkage method*, dan *Average Linkage method*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah pada penelitian ini, maka penelitian ini bertujuan untuk:

- 1) Bagaimana hasil pengelompokan nasabah KPR Bank X berdasarkan penilaian variabel 6 K menggunakan kombinasi antara metode pengukuran jarak *Euclidean distance*, *City-Block (Manhattan) distance*, dan *Mahalanobis distance* dengan metode pengelompokan *Ward's method*, *Complete Linkage method*, dan *Average Linkage method*?
- 2) Manakah metode pengukuran jarak *Euclidean distance*, *City-Block (Manhattan) distance*, dan *Mahalanobis distance* dengan metode pengelompokan *Ward's method*, *Complete Linkage method*, dan *Average Linkage method* yang memberikan hasil terbaik dalam pengelompokkan nasabah KPR Bank X.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- 1) Mengetahui pengelompokan nasabah KPR Bank X dengan menggunakan analisis *Cluster*
- 2) Hasil dari penelitian ini akan menunjukkan metode pengukuran kesamaan dan metode pengelompokan paling optimal, sehingga dapat digunakan menjadi acuan bagi penelitian selanjutnya.



- 3) Memberikan informasi kepada pihak perbankan untuk melakukan evaluasi dan strategi dalam pemberian kredit khususnya KPR

1.5. Batasan Masalah

Batasan masalah pada penelitian ini adalah:

- 1) Data yang digunakan adalah data primer, yaitu data penilaian 6 K pada nasabah KPR Bank X tahun 2021.
- 2) Metode pengukuran jarak yang digunakan adalah *Euclidean distance*, *City-Block (Manhattan) distance*, dan *Mahalanobis distance*.
- 3) Metode pengelompokan yang digunakan adalah *Ward's method*, *Complete Linkage method*, dan *Average Linkage method*.

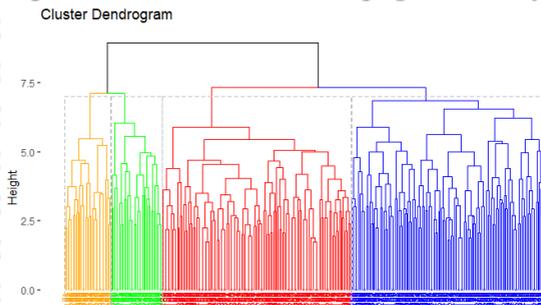
BAB II TINJAUAN PUSTAKA

2.1. Analisis Cluster

Analisis *cluster* (kelompok) merupakan salah satu analisis multivariat yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu dan dapat digabungkan dengan kelompok objek lainnya sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek-objek yang berada pada kelompok yang berbeda. Ciri-ciri suatu *cluster* yang baik adalah sebagai berikut:

1. *Homogenitas Internal*: Kesamaan antar anggota dalam satu *cluster*.
2. *Heterogenitas External*: Perbedaan antara *cluster* yang satu dengan *cluster* yang lain.

Pada dasarnya ada dua metode pada analisis *cluster*, yaitu metode hierarki dan metode non-hierarki. Metode hierarki merupakan metode analisis *cluster* secara bertahap. Banyaknya *cluster* yang terbentuk adalah $N - 1$ dengan N adalah banyaknya objek penelitian sehingga akan menghasilkan seluruh kemungkinan banyaknya *cluster* yang terbentuk. Metode ini akan menghasilkan sebuah diagram pohon (dendrogram) yang menggambarkan tahapan proses terbentuknya *cluster*. Metode hierarki dapat dibagi menjadi dua, yaitu proses *agglomerative* (penggabungan) dan proses *divisive* (pembagian/pemisahan).



Gambar 2.1 Dendrogram cluster hierarki



Proses *agglomerative* akan menjadikan setiap objek menjadi sebuah *cluster* sehingga jumlah *cluster* diawal akan sama dengan jumlah objek. Selanjutnya setiap *cluster* dengan tingkat kesamaan tinggi akan digabungkan menjadi sebuah *cluster* baru. Tahap ini diulang hingga hanya tersisa satu *cluster* yang berisi seluruh objek penelitian. Sedangkan metode *divisive* secara singkat merupakan kebalikan dari metode *agglomerative*. Seluruh objek yang ada akan digabungkan menjadi satu *cluster*, lalu objek dengan tingkat perbedaan yang tinggi akan dipisahkan dan menjadi *cluster* baru. Langkah tersebut dilakukan secara terus-menerus hingga setiap *cluster* hanya beranggotakan satu objek. Metode yang paling umum digunakan adalah metode hierarki *agglomerative*.

Berbeda dengan metode hierarki yang akan menghasilkan beberapa kemungkinan banyaknya *cluster* yang terbentuk, banyaknya *cluster* pada metode non-hierarki sudah ditentukan sejak awal, kemudian objek akan dimasukkan ke dalam *cluster-cluster* yang ada. Menurut Rencher (2002), ada tiga metode non-hierarki, yaitu *partitioning*, *mixture of distribution*, dan *density estimation*. Metode yang paling sering digunakan adalah metode *partitioning* dengan metode *K-Means*. Penelitian ini akan berfokus pada penggunaan analisis *cluster* hierarki *agglomerative*.

2.2. Metode Pengukuran Jarak

Pada metode hierarki *agglomerative*, dibutuhkan sebuah metode pengukuran kesamaan untuk menentukan objek yang akan digabungkan menjadi sebuah *cluster* baru. Menurut Hair, dkk. (2010), terdapat tiga metode pengukuran kesamaan, yaitu *correlational measures*, *distance measures*, dan *association measures*. Metode pengukuran kesamaan yang paling sering digunakan adalah *distance measures* (pengukuran jarak). Metode pengukuran jarak adalah metode yang dilakukan dengan mengukur jarak antar objek penelitian. Semakin kecil jarak antar objek, maka semakin homogen objek tersebut, begitupun sebaliknya. Menurut Hair, dkk. (2010), terdapat beberapa metode pengukuran jarak, antara lain *Euclidean distance*, *City-Block (Manhattan) distance*, dan *Mahalanobis distance*.



2.2.1. Euclidean Distance

Euclidean distance merupakan metode pengukuran yang paling umum digunakan. Biasanya dapat juga disebut sebagai jarak garis lurus. *Euclidean distance* antara dua titik objek dapat dianggap sebagai panjang sisi miring dari sebuah segitiga siku-siku. Metode ini membutuhkan data yang tidak memiliki korelasi antar variabel. Metode *Euclidean distance* ini mengukur jumlah kuadrat perbedaan nilai antara dua objek dengan menggunakan persamaan (2.1).

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

dimana:

$d(x_i, x_j)$: jarak antara objek ke- i dan ke- j

x_{ik} : data dari objek ke- i pada variabel ke- k

x_{jk} : data dari objek ke- j pada variabel ke- k

i, j : 1, 2, ..., n

k : 1, 2, ..., p

2.2.2. City-Block (Manhattan) Distance

Ukuran kedekatan menggunakan *Manhattan distance* adalah jarak antara dua objek yang merupakan jumlah perbedaan mutlak. Perhitungan jarak ini adalah yang paling mudah, tetapi dapat menghasilkan *cluster* yang tidak valid jika variabel yang digunakan memiliki korelasi yang tinggi. Jika *Euclidean distance* merupakan panjang sisi miring, maka *Manhattan distance* merupakan panjang kedua sisi tegak lurus dari segitiga siku-siku. Rumus jarak dapat ditulis seperti persamaan (2.2).

$$d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2.2)$$

dimana:

$d(x_i, x_j)$: jarak antara objek ke- i dan ke- j

x_{ik} : data dari objek ke- i pada variabel ke- k



x_{jk} : data dari objek ke- j pada variabel ke- k

i, j : 1, 2, ..., n $i \neq j$

k : 1, 2, ..., p

2.2.3. Mahalanobis Distance

Menurut Hair, dkk. (2010), *Mahalanobis distance* merupakan pengukuran jarak yang menyamaratakan seluruh objek dengan memperhatikan korelasi antar variabel tanpa membedakan beban tiap variabel. Menurut Seber (2004), *Mahalanobis distance* tidak hanya dapat mengatasi masalah perbedaan skala dalam data, tetapi juga mempertimbangkan pengaruh korelasi antar variabel. Pada saat variabel-variabel tidak saling berkorelasi, *Mahalanobis distance* dapat dianggap sama dengan *Euclidean distance* yang distandarisasi. *Mahalanobis distance* antar dua objek dapat dihitung menggunakan persamaan (2.3).

$$D_{ij}^2 = \frac{1}{(1-r^2)} \left[\frac{(x_{i1} - x_{j1})^2}{S_1^2} - 2r \frac{(x_{i1} - x_{j1})(x_{i2} - x_{j2})}{S_1 S_2} + \frac{(x_{i2} - x_{j2})^2}{S_2^2} \right] \quad (2.3)$$

dimana:

D_{ij}^2 : Jarak pengamatan ke- i dan ke- j

S_1^2 : varian untuk variabel ke-1

S_2^2 : varian untuk variabel ke-2

$S_1 S_2$: kovarian variabel ke-1 dan ke-2

2.3. Metode Pengelompokan

Jika pada saat sebuah *cluster* memiliki anggota lebih dari satu data, maka metode pengukuran jarak tidak dapat digunakan sehingga dibutuhkan sebuah metode untuk menggabungkan *cluster* yang memiliki anggota lebih dari satu. Menurut Mattjik, dkk. (2002), metode pengelompokan pada hirarki *agglomerative* digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya. Di samping itu, metode ini digunakan apabila pada awal analisis kurang adanya informasi tentang karakteristik objek-objek yang akan dikelompokkan sehingga banyaknya kelompok yang

diinginkan belum diketahui. Menurut Johnson dan Wichern (2007), ada beberapa metode pengelompokan pada analisis *cluster* hierarki, antara lain *Complete Linkage method*, *Average Linkage method*, dan *Ward method*.

2.3.1. Metode Ward

Menurut Johnson dan Wichern (2002), metode Ward merupakan suatu metode hierarki yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Jarak antara dua *cluster* yang terbentuk menggunakan metode Ward adalah *sum of square* diantara dua *cluster* tersebut. Metode Ward menggabungkan dua objek yang memiliki nilai *Sum of Squares Error* (SSE) terkecil diantara kemungkinan yang ada. Tujuan metode Ward meminimumkan ragam dalam *cluster* dan memaksimalkan ragam antar *cluster* lainnya (Supranto, 2004)

Proses pengelompokan pada metode Ward didasarkan pada kriteria SSE dengan ukuran kehomogenan antara dua objek berdasarkan jumlah kuadrat kesalahan. *Cluster* yang hanya berisi satu objek nilai SSE nya adalah nol. SSE hanya dapat dihitung jika *cluster* memiliki lebih dari satu objek menggunakan persamaan (2.5).

$$SSE = \sum_{i=1}^N (\bar{x}_i - \bar{x})^2 (\bar{x}_i - \bar{x})^2 \quad (2.5)$$

dimana:

\bar{x}_i : vektor kolom dari nilai objek ke- i , di mana $i = 1, 2, 3, \dots, n$

\bar{x} : vektor kolom rata-rata nilai seluruh objek

n : banyak objek dalam *cluster* yang terbentuk

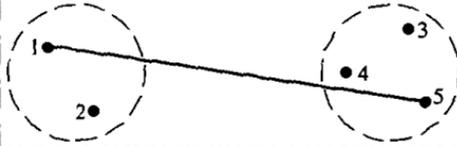
2.3.2. Complete Linkage Method

Jika dalam metode *Single Linkage* pengelompokan dilakukan berdasarkan jarak terdekat, maka dalam metode *Complete Linkage* pengelompokan dilakukan berdasarkan jarak terjauh dari semua *cluster* yang ada menggunakan persamaan (2.6).

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (2.6)$$



Nilai d_{uv} dan d_{vw} merupakan jarak antara *cluster* u dan w serta *cluster* v dan w .



Gambar 2. 2. Complete Linkage Method

Sumber: Johnson dan Wichern (2007)

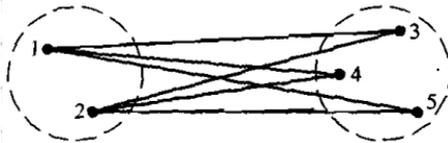
Metode *Complete Linkage* akan memilih dua objek dengan jarak paling jauh dan menggabungkannya menjadi sebuah *cluster*. Pada Gambar 2.3, akan terbentuk *cluster* d_{15} .

2.3.3. Average Linkage Method

Average Linkage Method menghitung rata-rata jarak dari seluruh objek di sebuah *cluster* dengan seluruh objek di *cluster* lain. Kemudian menggabungkan dua *cluster* dengan rata-rata jarak terkecil seperti pada persamaan (2.7).

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w} \quad (2.7)$$

Dengan d_{ik} merupakan jarak antar objek i dalam *cluster* uv dan objek k dalam *cluster* w . Sedangkan $N_{(uv)}$ dan N_w merupakan jumlah objek dalam *cluster* uv dan w .



Gambar 2. 3. Average Linkage Method

Sumber: Johnson dan Wichern (2007)

Berbeda dengan *Average Linkage* yang hanya melihat jarak antara dua objek, metode *Average Linkage* akan menghitung rata-rata



dari seluruh jarak yang ada pada dua *cluster*, lalu menggabungkan dua *cluster* yang memiliki rata-rata jarak terkecil.

2.4. Pemeriksaan Multikolinearitas

Menurut Ghozali (2018), tujuan uji multikolinearitas adalah untuk menguji apakah ditemukan adanya korelasi antar variabel. Multikolinearitas adalah keadaan ditemukannya korelasi yang sempurna atau mendekati sempurna antar variabel independen (Priyatno, 2012). Apabila variabel-variabel yang digunakan berkorelasi secara sempurna maka variabel-variabel tersebut akan mendapatkan bobot dua kali lebih besar daripada variabel lain. Metode pengukuran jarak menggunakan *Euclidean distance* dan *Manhattan distance* mengharuskan data terbebas dari korelasi antar variabel.

2.4.1. Variance Inflation Factor (VIF)

Variance Inflation Factor adalah alat untuk mengukur seberapa besar varians meningkat. Multikolinearitas terindikasi apabila nilai $VIF_h > 10$ dan salah satu cara yang dapat dilakukan jika terjadi multikolinearitas adalah dengan mengeluarkan variabel h dari model (Hair dkk., 2014). VIF dapat dihitung menggunakan persamaan (2.8).

$$VIF = \frac{1}{1 - R^2} \quad (2.8)$$

$$R^2 = \frac{\left[n \sum_{i=1}^n X_1 X_2 - \sum_{i=1}^n X_1 \sum_{i=1}^n X_2 \right]^2}{\left[n \sum_{i=1}^n (X_1)^2 - \left(\sum_{i=1}^n X_1 \right)^2 \right] \left[n \sum_{i=1}^n (X_2)^2 - \left(\sum_{i=1}^n X_2 \right)^2 \right]} \quad (2.9)$$

Jika data terbukti tidak memiliki korelasi antar variabel maka *Euclidean distance* dan *Manhattan distance* dapat digunakan untuk analisis *cluster*.

2.4.2. Analisis Komponen Utama (AKU)

Menurut Yamin (2011), AKU merupakan teknik statistik yang bertujuan untuk menyederhanakan variabel yang diamati dengan



mereduksi dimensinya (disebut juga sebagai teknik pereduksian data). Prinsip utama dalam AKU adalah terdapat korelasi di antara variabel. AKU digunakan untuk menjelaskan struktur matriks varians-kovarians dari suatu data set variabel melalui kombinasi linier dari variabel-variabel. AKU akan mentransformasi variabel tersebut ke dalam komponen atau variabel baru yang tidak berkorelasi sama sekali. Setelah terbentuk beberapa komponen hasil AKU yang bebas multikolinearitas, komponen-komponen tersebut menjadi variabel estimasi baru yang akan digunakan untuk analisis *cluster*. Kegunaan analisis komponen utama antara lain:

1. Identifikasi variabel baru yang mendasari data multivariabel.
2. Mengurangi banyaknya dimensi himpunan variabel yang biasanya terdiri atas variabel yang banyak dan saling berkorelasi menjadi variabel-variabel baru yang tidak berkorelasi dengan mempertahankan sebanyak mungkin keragaman (informasi) dalam himpunan data tersebut.
3. Menghilangkan variabel-variabel asal yang mempunyai sumbangan informasi yang relatif kecil.

Pada analisis komponen utama terdapat matriks masukan, yaitu matriks korelasi yang disajikan pada persamaan (2.10).

$$\hat{\rho} = \frac{\sigma_{xy}}{\sqrt{(\hat{\sigma}_x \hat{\sigma}_y)}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$\mathbf{r} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix} \quad (2.10)$$

Setelah nenbentuk natriks korelasi sampel dapat diperoleh nilai eigen ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$). Nilai eigen diperoleh dengan cara menyelesaikan persamaan karakteristik dari matriks korelasi seperti pada persamaan (2.11).

$$|\mathbf{r} - \lambda \mathbf{I}| = 0 \quad (2.11)$$



Menurut Solimun, dkk. (2017), terdapat tiga metode yang dapat digunakan untuk menentukan banyaknya komponen utama, salah satunya adalah dengan proporsi kumulatif dari keragaman total. Rumus proporsi keragaman dari k nilai eigen dapat dilihat pada persamaan (2.12).

$$\frac{\sum_{i=1}^k \lambda_i}{p}, k = 1, \dots, p \quad (2.12)$$

Secara teoritis tidak ada petunjuk yang jelas untuk menentukan kriteria yang paling tepat digunakan, sehingga dapat dipilih salah satu bila keduanya tidak konsisten. Pemilihan komponen yang terbentuk didasarkan atas pertimbangan pragmatis, yaitu disesuaikan dengan hasil analisis, kemudahan interpretasi dan kegunaannya di dalam bidang yang dikaji (Astutik, dkk. 2018). Pada penelitian ini penentuan komponen utama yang terbentuk adalah komponen yang pertama dengan rumus untuk menghitung bobot komponen pada persamaan (2.13).

$$(\mathbf{X}\mathbf{X} - \lambda_i \mathbf{I})\mathbf{b}_j = 0 \quad (2.13)$$

Setelah mendapatkan bobot komponen, selanjutnya dapat membentuk persamaan komponen utama linier. Misal terdapat p variabel X , yaitu X_1, X_2, \dots, X_p kombinasi linier komponen utama yang terbentuk seperti pada persamaan (2.14).

$$\begin{aligned} PC_1 &= b_{11}X_1 + b_{12}X_2 + \dots + b_{1p}X_p + \varepsilon_1 \\ PC_2 &= b_{21}X_1 + b_{22}X_2 + \dots + b_{2p}X_p + \varepsilon_2 \\ &\vdots \\ PC_p &= b_{p1}X_1 + b_{p2}X_2 + \dots + b_{pp}X_p + \varepsilon_p \end{aligned} \quad (2.14)$$

dimana:

PC_p : banyak dimensi yang terbentuk sebanyak p

b_{pp} : bobot komponen

X_p : banyak indikator sebanyak $p, j=1, 2, \dots, p$



Apabila komponen utama telah diperoleh dapat menghitung skor komponen dari setiap individu yang akan digunakan sebagai analisis yang lebih lanjut. Menurut Astutik, dkk. (2018), vektor skor individu ke- i adalah seperti $(X_{1i}, X_{2i}, \dots, X_{pi})$ atau dapat dituliskan dalam bentuk matriks data pada persamaan (2.15).

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \quad (2.15)$$

Maka skor komponen dari individu ke- i pada komponen utama K_j yang dihasilkan dari matriks korelasi dapat dilihat pada persamaan (2.16).

$$SK_{ij} = (b_{1j}, b_{2j}, \dots, b_{pj}) \begin{bmatrix} X_{i1} - \bar{X}_1 \\ X_{i2} - \bar{X}_2 \\ \vdots \\ X_{ip} - \bar{X}_p \end{bmatrix} \quad (2.16)$$

Persamaan (2.16) dapat dituliskan sebagaimana pada persamaan (2.17).

$$SK_{ij} = \mathbf{b}'_j (\mathbf{X}_i - \bar{\mathbf{X}}) \quad (2.17)$$

dimana:

SK_{ij} : skor komponen ke- j dari individu ke- i

\mathbf{b}'_j : vektor pembobot komponen utama ke- j

\mathbf{X}_i : vektor data individu ke- i

$\bar{\mathbf{X}}$: vektor nilai rata-rata dari peubah asal

Jika komponen utama dihasilkan dari matriks korelasi \mathbf{r} , maka rumus skor komponen utama dari individu ke- i dapat ditulis dalam persamaan (2.18).

$$SK_{ij} = \mathbf{z}'_i \mathbf{b}_j \quad (2.18)$$

Keterangan:

\mathbf{z}_i : vektor data individu ke- i yang telah dibakukan



2.5. Indeks Validitas Cluster

Menurut Hair, dkk. (2006), validitas *cluster* dapat digunakan untuk mengetahui banyaknya cluster yang telah terbentuk dapat menjelaskan dan mewakili populasi secara umum atau tidak. Validitas *cluster* dapat digunakan untuk membantu memecahkan permasalahan utama dalam analisis *cluster*, yaitu menentukan banyaknya *cluster* (kelompok) yang optimum. *Cluster* optimum yang dimaksud adalah *cluster* yang mempunyai jarak yang padat antar objek, tetapi memiliki jarak yang jauh dari *cluster* lainnya (Dubes dan Jain, 1988). Pemeriksaan validitas *cluster* mampu mengevaluasi kebaikan dari hasil analisis *cluster* secara kuantitatif sehingga mampu menghasilkan *cluster* optimum.

2.5.1. Indeks Dunn (D)

Indeks validasi *Dunn* dilambangkan dengan D yang dihitung menggunakan persamaan (2.19).

$$D = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, i \neq j} \left\{ \frac{d(c_i, c_j)}{\max_{1 \leq k \leq n} (d^*(c_k))} \right\} \right\} \quad (2.19)$$

dimana:

$d(c_i, c_j)$: jarak antar kelompok c_i dan c_j

$d^*(c_k)$: jarak dalam kelompok c_k

Indeks *Dunn* akan menunjukkan kedekatan objek-objek di dalam *cluster* dan antar *cluster* yang terbentuk. Antar objek didalam *cluster* harus memiliki jarak minimum sehingga diameter *cluster* akan berukuran kecil, tetapi memiliki jarak maksimum dengan *cluster* lain. Semakin besar nilai indeks *Dunn* mengindikasi bahwa banyak *cluster* yang terbentuk dianggap tepat (Kovacs dkk., 2005).

2.5.2. Indeks Davies-Bouldin (DB)

Rumus untuk indeks *Davies-Bouldin* dapat dituliskan sesuai dengan persamaan (2.20).

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left[\frac{d^*(c_i) + d^*(c_j)}{d(c_i, c_j)} \right] \quad (2.20)$$



dimana:

n : jumlah kelompok

$d(c_i, c_j)$: jarak antar kelompok c_i dan c_j

$d^*(c_k)$: jarak dalam kelompok c_k

Nilai indeks *Davies-Bouldin* yang semakin kecil menunjukkan banyaknya *cluster* yang terbentuk telah baik (Su, 2003).

2.5.3. Indeks *Global Silhouette* (*Gsu*)

Penghitungan indeks *Global Silhouette* dipisah menjadi dua tahap. Rumus penghitungan indeks *Silhouette* $S(i)$ dapat dilihat pada persamaan (2.21).

$$S(i) = \frac{(b(i) - a(i))}{\max\{a(i), b(i)\}} \quad (2.21)$$

dimana:

$a(i)$: rata-rata perbedaan dari i -objek dengan semua objek lain di dalam kelompok yang sama.

$b(i)$: nilai minimum dari rata-rata perbedaan dari i objek dengan semua objek pada kelompok lain.

Nilai indeks *Silhouette* yang semakin besar menunjukkan bahwa jumlah *cluster* yang terbentuk telah optimal. Rata-rata $S(i)$ dari seluruh objek dalam suatu *cluster* menunjukkan kemiripan objek dalam *cluster* tersebut dan ketepatan objek setelah dikelompokan. Rata-rata indeks $S(i)$ dapat disebut indeks *Global Silhouette* dengan rumus pada persamaan (2.22).

$$GS_u = \frac{1}{n} \sum_{i=1}^n S(i) \quad (2.22)$$

dimana:

$S(i)$: *Silhouette* kelompok ke- i

n : jumlah kelompok



Nilai yang paling besar dari indeks *Global Silhouette* menunjukkan jumlah kelompok terbaik yang kemudian diambil sebagai kelompok optimum.

2.5.4. Indeks Gabungan

Ketiga indeks validitas *cluster* yang digunakan dapat memberikan hasil jumlah *cluster* optimal yang berbeda. Alternatif untuk memilih jumlah *cluster* optimal menggunakan indeks gabungan, yaitu dengan mengombinasikan indeks validitas *cluster* kemudian dipilih jumlah *cluster* optimal saat indeks tersebut memiliki kombinasi paling kecil (Azuaje dan Bolshakova, 2001). Caranya dengan menghitung ketiga indeks validitas *cluster*, lalu memberi *ranking* pada tiap jumlah *cluster* yang mungkin terbentuk pada masing-masing indeks. Jumlah *cluster* dianggap baik jika memiliki *ranking* terkecil.

2.6. Penentuan Kebaikan Metode

Pada perbandingan penggunaan beberapa metode pada analisis *cluster*, akan dibutuhkan sebuah perhitungan untuk menentukan metode mana yang memberikan hasil paling optimal. Simpangan baku dapat digunakan untuk mengetahui kehomogenan antar *cluster*. *Cluster* yang terbentuk dikatakan baik apabila mempunyai nilai simpangan baku dalam *cluster* (S_w) yang minimum dan simpangan baku antar *cluster* (S_B) yang maksimum (Hardle dan Simar, 2003).

Rumus simpangan baku dalam *cluster* (S_w) dapat dilihat pada persamaan (2.23).

$$S_w = \frac{\sum_{k=1}^K S_k}{K} \quad (2.23)$$

dimana:

S_k : simpangan baku *cluster* ke - k

K : banyaknya *cluster* yang terbentuk



Sedangkan untuk simpangan baku antar *cluster* (S_B) dapat menggunakan persamaan (2.24).

$$S_B = \sqrt{\frac{\sum_{k=1}^K (\bar{X}_k - \bar{X})^2}{(K-1)}} \quad (2.24)$$

dimana:

\bar{X}_k : rata-rata *cluster* ke - k

\bar{X} : rata-rata seluruh *cluster*

K : banyaknya *cluster* yang terbentuk

Nilai rasio simpangan baku didapatkan dari nilai (S_w) dan (S_B). Metode dengan nilai rasio simpangan baku terbesar dapat dianggap sebagai metode paling optimal. Semakin besar nilai rasio simpangan baku suatu metode, maka semakin baik metode yang digunakan. Rasio simpangan baku dapat menggunakan persamaan (2.25).

$$\text{rasio simpangan baku} = \frac{S_B}{S_w} \quad (2.25)$$

dimana

S_w : simpangan baku dalam *cluster*

S_B : simpangan baku antar *cluster*

2.7. Variabel dan Data Penelitian

Menurut Yitnosumarto dalam Solimun, dkk. (2017), data merupakan kumpulan angka, keadaan, fakta pendapat dan lainnya yang merupakan hasil pengamatan, pengukuran atau pencacahan terhadap variabel dari suatu objek penelitian. Variabel adalah karakteristik yang diketahui dari objek yang diamati dan dapat memberikan penjelasan terhadap objek (Suwarno, 2005). Berdasarkan sumbernya, data dapat dibagi menjadi dua, yaitu data yang didapatkan



secara langsung dari sumbernya (data primer) dan data yang didapatkan bukan dari sumbernya secara langsung (data sekunder).

Jenis data berdasarkan skala ukurnya dibedakan menjadi data nominal, ordinal, interval dan rasio. Hal ini dibutuhkan untuk menentukan metode analisis data statistika yang akan digunakan. Data interval dan rasio termasuk dalam jenis data numerik, sedangkan data nominal dan ordinal bersifat kualitatif, sehingga harus diubah menjadi bentuk numerik dengan cara pemberian skor (*scoring*).

Menurut Solimun, dkk. (2017), penelitian di bidang sosial sering menganalisa atribut psikologis yang merupakan variabel yang tidak dapat diukur secara langsung atau dapat disebut sebagai variabel laten (*unobservable*), sehingga pengukuran variabel laten memiliki peran yang penting, yaitu dengan menggunakan skala sikap.

2.1.1. Skala Sikap Instrumen Penelitian

Menurut Riduwan (2015), terdapat beberapa skala sikap, seperti berikut.

1. Skala *Likert*

Skala *Likert* digunakan untuk mengukur sikap, pendapat dan persepsi seseorang tentang kejadian atau keadaan, skala ini memiliki jarak sikap, misalnya sangat tidak setuju sampai dengan sangat setuju.

2. Skala *Guttman*

Skala *Guttman* digunakan untuk mengukur dimensi dari sebuah variabel yang bersifat jelas, tegas dan konsisten. Skala ini menghasilkan data dikotomi, misalnya setuju dan tidak setuju.

3. Skala *Diferensial Semantik*

Pada skala *Diferensial Semantik* menggunakan respon/jawaban berupa sebuah garis kontinu dengan nilai sangat tidak setuju berada di ujung paling kiri, dan sangat setuju berada di ujung paling kanan.

4. Skala *Stapel*

Skala *Stapel* digunakan untuk mengukur sikap yang diberi nilai negatif dan positif.

5. Skala *Thurstone*



Pada skala *Thurstone*, responden akan diminta untuk memilih pernyataan yang disetujui dari beberapa pernyataan yang menyajikan data berbeda-beda.

2.8. Metode Pengukuran Variabel

Menurut Solimun, dkk. (2017), berikut ini adalah beberapa cara untuk mendapatkan data variabel laten:

1. Metode Skor Total

Metode ini menggunakan jumlah skor semua indikator pada masing-masing variabel sehingga diperoleh data skor total.

2. Metode Skor Rata-Rata

Metode ini menggunakan skor rata-rata semua indikator pada setiap variabel sehingga diperoleh data skor rata-rata yang merupakan data variabel laten.

3. Metode *Rescoring*

Metode *rescoring* mengubah total skor menjadi skor awal yang digunakan, yaitu 1 sampai 5, kemudian menghitung data variabel laten.

2.9. Pemeriksaan Instrumen Penelitian

Salah satu cara untuk mendapatkan data yang dapat digunakan untuk penelitian adalah dari hasil kuesioner. Sebuah kuesioner akan menghasilkan data yang akurat apabila telah diperiksa tingkat validitas dan reliabilitas hingga dinyatakan valid dan reliabel. Berikut ini merupakan penjelasan mengenai pemeriksaan validitas dan reliabilitas.

2.9.1. Pemeriksaan Validitas

Menurut Solimun, dkk. (2017), validitas kuesioner dapat menunjukkan sebaik apa sebuah kuesioner yang digunakan untuk mengukur objek yang ingin diukur sesuai dengan kondisi sebenarnya. Terdapat beberapa jenis validitas kuesioner, antara lain:

1. Validitas Isi

Validitas isi dapat ditentukan oleh ketepatan definisi operasional berdasarkan teori yang relevan.

2. Validitas Kriteria



Validitas kriteria dapat dilakukan dengan menghitung koefisien korelasi antara skor masing-masing item dengan skor total terkoreksi.

3. Validitas Unidimensional

Kriteria yang sering digunakan adalah melihat keragaman kumulatif dan jika nilainya sekitar 70%, maka kuesioner memenuhi validitas.

4. Validitas Konvergen

Validitas konvergen dapat diukur dengan muatan faktor untuk model indikator reflektif dan bobot komponen untuk model indikator formatif. Indikator dikatakan signifikan apabila muatan faktor dan bobot komponen $\geq 0,30$.

5. Validitas Diskriminan

Validitas diskriminan dapat diukur dengan menggunakan perbandingan antar *loading* dengan *crossloading factor*. Dapat dikatakan memenuhi validitas diskriminan jika nilai *loading* lebih besar dari nilai *crossloading*.

Pada penelitian ini, pengujian validitas yang akan digunakan adalah *corrected item total correlation*, sehingga dapat diperoleh informasi yang lebih akurat. Menurut Kline (2000), koreksi terhadap nilai koefisien korelasi yang *overestimate* ditunjukkan pada persamaan (2.14).

$$r_{i(T-i)} = \frac{r_{ii} s_T - s_i}{\sqrt{s_i^2 + s_T^2 - 2s_i s_T r_{ii}}} \quad (2.27)$$

dimana:

$r_{i(T-i)}$: koefisien korelasi dari item ke- i dengan total skor semua item kecuali item ke- i

r_{ii} : koefisien korelasi dari item dengan total skor

s_T : standar deviasi dari total skor

s_i : standar deviasi dari item ke- i

Menurut Masrun dalam Solimun (2010), jika nilai *corrected item total correlation* item pertanyaan bernilai positif dan lebih dari 0,30, maka item dalam instrumen penelitian tersebut dapat dinyatakan valid. Apabila seluruh instrumen penelitian telah dinyatakan valid,



selanjutnya dapat dilakukan pemeriksaan reliabilitas instrumen penelitian.

2.9.2. Pemeriksaan Reliabilitas

Solimun, dkk. (2017) menjelaskan bahwa realibilitas adalah ukuran yang menunjukkan sejauh mana sebuah kuesioner mampu mengukur sebuah variabel secara konsisten. Menurut Malhotra dalam Solimun, dkk. (2017), uji reliabilitas dapat menggunakan metode *test re-test* dan *internal consistency* dengan cara memeriksa koefisien *alpha cronbach*. Instrumen penelitian dapat dinyatakan reliabel jika nilai $\alpha \geq 0,60$ yang dihitung menggunakan persamaan (2.15).

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right) \quad (2.28)$$

dimana:

α : koefisien reliabilitas *alpha cronbach*

k : banyaknya item pernyataan

S_i^2 : ragam skor item ke- i

S_t^2 : ragam skor total item

2.10. Variabel Penelitian

Dalam upaya untuk meminimalisir risiko yang disebabkan oleh calon nasabah, bank dapat melakukan identifikasi kepada nasabah berdasarkan prinsip 6 K, yaitu karakter, kapasitas, kapital, kekayaan, kondisi, dan kepatuhan (Astiko dan Sunardi, 1996). Hal ini diperlukan untuk memberikan besaran kredit yang sesuai dengan kebutuhan calon nasabah dan dapat dipastikan bahwa nasabah memiliki kemampuan untuk melunasinya sesuai dengan batas waktu yang disepakati.

2.10.1. Karakter

Sebelum memberikan kredit, pihak bank harus mengerti karakter dari calon nasabah seperti kejujuran, integritas dan iktikad baik. Penilaian bank terhadap karakter calon nasabah dibutuhkan untuk mengetahui bahwa calon nasabah tersebut memiliki sifat-sifat positif dan memiliki catatan baik pada Sistem Informasi Nasabah





(SID) di Bank Indonesia (BI), sehingga tidak akan menimbulkan masalah bagi bank dan dapat melunasi kredit yang diambil tepat waktu sesuai dengan perjanjian.

2.10.2. Kapasitas

Penilaian bank atas kemampuan calon nasabah untuk membayar pinjaman beserta bunganya, ditinjau dari kemampuan manajemen keuangan calon nasabah dan prospek dari usaha yang akan dilakukan, termasuk kemampuan calon nasabah untuk menjalankan usahanya agar dapat mendapatkan keuntungan sesuai dengan yang diharapkan, sehingga bank yakin uang yang dipinjam akan digunakan dengan baik.

2.10.3. Kapital

Capital adalah penilaian bank atas kondisi keuangan calon nasabah secara keseluruhan, termasuk catatan keuangan calon nasabah di masa lalu maupun proyeksi di masa depan, sehingga bank akan mengetahui modal yang dimiliki oleh calon nasabah. Semakin besar modal yang dimiliki, maka semakin besar pula kemampuan calon nasabah untuk membayar pinjamannya.

2.10.4. Kekayaan

Penilaian bank terhadap aset kekayaan yang dimiliki oleh calon nasabah yang dapat dijadikan sebagai jaminan. Aset tersebut dapat berupa benda berwujud maupun tidak berwujud yang akan diserahkan kepada bank untuk menjamin pelunasan utang nasabah. Jika nasabah tidak dapat melakukan pelunasan sesuai batas waktu yang sudah disepakati, maka bank dapat menyita aset yang sudah dijadikan jaminan. Besaran kredit berbanding lurus dengan banyaknya aset.

2.10.5. Kondisi

Kondisi yang dimaksudkan disini adalah kondisi politik, sosial, budaya, ekonomi maupun pasar baik di dalam negeri maupun di luar negeri pada saat ini maupun kemungkinan di masa yang akan datang. Hal ini dapat mempengaruhi secara langsung maupun tidak langsung kemampuan atau kelancaran usaha dari calon nasabah. Kondisi yang tidak stabil akan memperkecil kemampuan calon nasabah untuk membayar kewajibannya.



2.10.6. Kepatuhan

Perilaku patuh membayar dapat diartikan sebagai kemampuan dan tanggung jawab calon nasabah kredit untuk membayar angsuran yang telah disetujui secara tepat waktu sesuai dengan perjanjian.

Perilaku ini dapat dipengaruhi oleh beberapa faktor, contohnya ketepatan waktu pembayaran angsuran, akurasi data yang dimiliki oleh pihak bank terkait dengan rekam jejak pembayaran yang dilakukan oleh calon nasabah dan sanksi yang akan diterima oleh nasabah jika melanggar kesepakatan yang sudah ditentukan.



BAB III METODE PENELITIAN

3.1. Sumber Data

Data yang digunakan adalah data primer tentang penilaian 6 K terhadap nasabah kredit pada Bank X dengan cara memberikan kuesioner dengan skala Likert. Pada data tersebut, terdapat catatan penilaian 6 K dari nasabah Kredit Pemilikan Rumah (KPR). Pengukuran variabel menggunakan metode skor rata-rata. Jumlah responden yang digunakan pada penelitian ini sebanyak 300 responden dari nasabah KPR pada Bank X. Struktur data dari penelitian ini dapat dilihat pada Tabel 3.1.

Tabel 3. 1. Struktur Data Penelitian

No.	Y_1	Y_2	Y_3	...	Y_k
1	$Y_{1,1}$	$Y_{2,1}$	$Y_{3,1}$	\vdots	$Y_{k,1}$
2	$Y_{1,2}$	$Y_{2,2}$	$Y_{3,2}$	\vdots	$Y_{k,2}$
3	$Y_{1,3}$	$Y_{2,3}$	$Y_{3,3}$	\vdots	$Y_{k,3}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	$Y_{1,n}$	$Y_{2,n}$	$Y_{3,n}$...	$Y_{k,n}$

keterangan:

$Y_{k,n}$: Nilai pada variabel penelitian ke- k dari objek ke- n

k : Banyaknya variabel

n : Banyak objek penelitian

3.2. Lokasi dan Waktu Penelitian

Lokasi yang digunakan untuk penelitian ini adalah Bank X di salah satu kota di Indonesia. Penelitian ini dilakukan pada bulan November 2021.

3.3. Populasi dan Sampel

Unit analisis dan unit sampel dalam penelitian ini adalah individu. Populasinya adalah seluruh kreditur KPR pada Bank X, sedangkan sampel yang digunakan adalah sebanyak 300 nasabah



skreditur KPR pada Bank X. Penelitian ini menggunakan kuesioner dengan teknik penarikan sampel *purposive sampling* yaitu pengambilan sampel secara sengaja sesuai persyaratan sampel yang diperlukan. Menurut Sugiono (2018), *purposive sampling* adalah pengambilan sampel dengan pertimbangan tertentu berdasarkan kriteria sampel yang dibutuhkan. Pada penelitian ini kriteria responden yang diperlukan adalah sebagai berikut.

1. Nasabah yang sedang menjadi kreditur KPR di Bank X.
2. Nasabah yang telah melewati setengah dari masa angsuran.

3.4. Variabel Penelitian

Penelitian ini menggunakan 6 variabel, yaitu karakter (Y_1), kapasitas (Y_2), kapital (Y_3), kekayaan (Y_4), kondisi (Y_5) dan kepatuhan (Y_6). Seluruh variabel yang digunakan memiliki skala interval.

3.5. Kisi-Kisi Instrumentasi

Kisi-kisi instrumentasi merupakan langkah pertama guna membuat kuesioner. Berikut adalah tabel kisi-kisi instrumen penelitian yang digunakan dalam penelitian:

Tabel 3. 2. Kisi-Kisi Instrumen Penelitian

Variabel Penelitian	Indikator	Item	Skala Data
Karakter (Y_1)	Itikad dan tanggung jawab ($Y_{1.1}$)	Secara rutin membayar cicilan KPR tepat waktu	Likert
		Berusaha agar tidak pernah terlambat dalam membayar angsuran KPR	Likert
	Sifat/ watak/ gaya hidup ($Y_{1.2}$)	Menghabiskan sebagian besar penghasilan untuk keperluan keluarga	Likert
	Komitmen pembayaran ($Y_{1.3}$)	Selalu menabung kelebihan penghasilan	Likert
		Merasa bersalah jika tidak membayar angsuran Kredit Pemilikan Rumah tepat waktu	Likert

Tabel 3.2. Lanjutan

	Komitmen pembayaran (Y _{1.3})	Merasa harus membayar angsuran tepat waktu karena kredit disetujui	Likert
	Pendapatan nasabah (Y _{2.1})	Memiliki penghasilan kurang dari standar UMR	Likert
		Pendapatan hanya berasal dari satu sumber	Likert
Kapasitas (Y ₂)	Kemampuan dalam membayar angsuran (Y _{2.2})	Mampu membayar angsuran secara berkala	Likert
		Keuntungan usaha dapat digunakan untuk membayar angsuran	Likert
	Kemampuan dalam menyelesaikan kredit tepat waktu (Y _{2.3})	Siap menerima sanksi jika tidak mampu membayar cicilan KPR	Likert
		Uang modal dapat kembali dalam jangka waktu kredit	Likert
Kapital (Y ₃)	Sumber penghasilan tetap (Y _{3.1})	Saya memiliki penghasilan yang tetap	Likert
		Saya memiliki pekerjaan yang tetap	Likert
	Memiliki bidang usaha lain sebagai sumber penghasilan (Y _{3.2})	Memiliki pekerjaan lain untuk membayar angsuran KPR	Likert
Memiliki usaha sendiri untuk meningkatkan penghasilan		Likert	
	Kemampuan mengelola keuangan (Y _{3.3})	Pengelolaan keuangan di keluarga sudah efisien	Likert
		Dapat menggunakan amanah keuangan secara baik dan benar	Likert

Tabel 3. 2. Lanjutan

Kekayaan (Y _{4.1})	Nilai jual barang jaminan (Y _{4.1})	Memiliki barang jaminan yang berharga untuk dijadikan jaminan jika saya tidak dapat melunasi KPR	Likert
		Nilai aset yang dimiliki melebihi plafond pinjaman kredit	Likert
	Jaminan bersifat fisik, atau non fisik (Y _{4.2})	Masih memiliki jaminan fisik jika saya gagal melunasi KPR	Likert
		Jika gagal melunasi KPR saya merelakan aset berharga saya untuk dilelang oleh pihak bank	Likert
	Kepemilikan barang jaminan dan keaslian dokumen (Y _{4.3})	Memiliki bukti keaslian aset berharga yang menjadi jaminan yaitu berupa sertifikat	Likert
		Seluruh aset yang saya miliki merupakan milik pribadi	Likert
Kondisi (Y _{5.1})	Pengembangan bisnis/usaha/ investasi (Y _{5.1})	Usaha yang dijalani banyak peminatnya sehingga penghasilan saya dapat digunakan untuk membayar KPR	Likert
		Mudah untuk mengembangkan usaha yang akan saya jalankan sehingga keuntungan akan terus bertambah	Likert



Tabel 3.2. Lanjutan

Variabel Penelitian	Indikator	Item	Skala Data
	Fluktuasi perekonomian ($Y_{5,2}$)	Pekerjaan yang dilakukan masih dapat bertahan meskipun ada ancaman krisis ekonomi	Likert
		Fluktuasi perekonomian akan berpengaruh terhadap usaha yang dijalankan	Likert
Kondisi (Y_5)	Kondisi sosial ekonomi/ problematika keluarga ($Y_{5,3}$)	Keluarga saya termasuk keluarga yang berkecukupan sehingga penghasilan yang saya terima masih mampu untuk membayar cicilan KPR tepat waktu	Likert
		Keluarga saya tidak memiliki hutang, sehingga saya dapat membayar cicilan Kredit Pemilikan Rumah tepat waktu	Likert
Kepatuhan (Y_6)	Keinginan Selalu Membayar Tepat Waktu ($Y_{6,1}$)	Keinginan untuk membayar cicilan Kredit Pemilikan Rumah tepat pada waktunya	Likert
		Keinginan membayar cicilan Kredit Pemilikan Rumah sebelum jatuh tempo	Likert

Tabel 3. 2. Lanjutan

Variabel Penelitian	Indikator	Item	Skala Data
Kepatuhan (Y ₆)	Akurasi data nasabah (Y _{6.2})	Saya selalu membayar cicilan Membayar sesuai dengan besaran angsuran	Likert
		Saya selalu melunasi kredit yang saya ambil	Likert
	Sanksi pembayaran kredit (Y _{6.3})	Sanksi dibuat untuk menumbuhkan kedisiplinan dari nasabah	Likert
		Sanksi digunakan sebagai efek jera bagi yang melanggar	Likert

3.6. Evaluasi Instrumen Penelitian

Pada umumnya, pemeriksaan instrumen penelitian disarankan minimal dilakukan kepada 30 responden karena distribusi skor untuk 30 responden diperkirakan akan mendekati distribusi normal yang berguna untuk penghitungan statistika (Solimun dkk., 2017). Pemeriksaan validitas dilakukan pada setiap item pertanyaan yang ada pada kuesioner. Instrumen dinyatakan valid jika nilai *corrected item total correlation* $\geq 0,3$. Jika nilainya kurang dari 0,3 maka item tersebut harus dihapuskan atau diganti. Hasil pemeriksaan validitas terhadap instrumen penelitian yang telah dibuat disajikan pada Tabel 3.3.

Tabel 3. 3. Uji Validitas Instrumen Penelitian

Variabel	Indikator	Item	<i>Corrected item total correlation</i>	Keterangan
Karakter (Y ₁)	Itikad dan tanggung jawab (Y _{1.1})	Y _{1.1.1}	0,314	Valid
		Y _{1.1.2}	0,324	Valid
	Sifat/ watak/ gaya hidup (Y _{1.2})	Y _{1.2.1}	0,400	Valid
		Y _{1.2.2}	0,378	Valid

Tabel 3. 3. Lanjutan

Variabel	Indikator	Item	Corrected item total correlation	Keterangan
	Komitmen pembayaran (Y _{1.3})	Y _{1.3.1}	0,429	Valid
		Y _{1.3.2}	0,391	Valid
Kapasitas (Y ₂)	Pendapatan nasabah (Y _{2.1})	Y _{2.1.1}	0,393	Valid
		Y _{2.1.2}	0,308	Valid
	Kemampuan dalam membayar angsuran (Y _{2.2})	Y _{2.2.1}	0,382	Valid
		Y _{2.2.2}	0,478	Valid
	Kemampuan dalam menyelesaikan kredit tepat waktu (Y _{2.3})	Y _{2.3.1}	0,451	Valid
		Y _{2.3.2}	0,435	Valid
Kapital (Y ₃)	Sumber penghasilan tetap (Y _{3.1})	Y _{3.1.1}	0,429	Valid
		Y _{3.1.2}	0,476	Valid
	Memiliki bidang usaha lain sebagai sumber penghasilan (Y _{3.2})	Y _{3.2.1}	0,503	Valid
		Y _{3.2.2}	0,531	Valid
Kemampuan mengelola keuangan (Y _{3.3})	Y _{3.3.1}	0,470	Valid	
	Y _{3.3.2}	0,458	Valid	
<i>Collateral</i> (Y ₄)	Nilai jual barang jaminan (Y _{4.1})	Y _{4.1.1}	0,463	Valid
		Y _{4.1.2}	0,372	Valid
	Jaminan bersifat fisik, atau non fisik (Y _{4.2})	Y _{4.2.1}	0,335	Valid
		Y _{4.2.2}	0,306	Valid
	Kepemilikan barang jaminan (Y _{4.3})	Y _{4.3.1}	0,406	Valid
		Y _{4.3.2}	0,371	Valid

Tabel 3. 3. Lanjutan

Variabel	Indikator	Item	<i>Corrected item total correlation</i>	Keterangan
Kondisi (Y ₅)	Pengembangan bisnis/usaha/ investasi (Y _{5.1})	Y _{5.1.1}	0,414	Valid
		Y _{5.1.2}	0,357	Valid
	Fluktuasi perekonomian (Y _{5.2})	Y _{5.2.1}	0,335	Valid
		Y _{5.2.2}	0,466	Valid
	Kondisi sosial ekonomi/ problematika keluarga (Y _{5.3})	Y _{5.3.1}	0,552	Valid
		Y _{5.3.2}	0,535	Valid
Kepatuhan (Y ₆)	Keinginan Selalu Membayar Tepat Waktu (Y _{6.1})	Y _{6.1.1}	0,503	Valid
		Y _{6.1.2}	0,452	Valid
	Akurasi data nasabah (Y _{6.2})	Y _{6.2.1}	0,335	Valid
		Y _{6.2.2}	0,498	Valid
	Sanksi pembayaran kredit (Y _{6.3})	Y _{6.3.1}	0,398	Valid
		Y _{6.3.2}	0,327	Valid

Tabel 3. 4. Pemeriksaan Reliabilitas Kuesioner

Variabel	Nilai <i>Cronbach Alpha</i>	Keterangan
Karakter (Y ₁)	0,734	Reliabel
Kapasitas (Y ₂)	0,675	Reliabel
Kapital (Y ₃)	0,734	Reliabel
Kekayaan (Y ₄)	0,644	Reliabel
Kondisi (Y ₅)	0,711	Reliabel
Kepatuhan (Y ₆)	0,681	Reliabel

Pada Tabel 3.3 dapat diketahui bahwa seluruh item memiliki nilai *corrected item total correlation* $\geq 0,30$, sehingga kuesioner yang



telah dibuat dinyatakan valid dan dapat dilanjutkan dengan uji reliabilitas dengan hasil tersaji pada Tabel 3.4.

Berdasarkan Tabel 3.4, seluruh variabel yang digunakan memiliki nilai *cronbach alpha* $\geq 0,60$, sehingga seluruh variabel dinyatakan reliabel. Pemeriksaan instrumen penelitian yang telah dilakukan menunjukkan bahwa seluruh instrumen yang telah dibuat dinyatakan valid dan reliabel sehingga dapat digunakan untuk mengumpulkan data dengan memberikan kuesioner kepada nasabah KPR Bank X.

3.7. Metode Penelitian

Langkah-langkah penelitian adalah sebagai berikut:

- a. Menentukan lokasi dan waktu penelitian yang telah dibahas pada subbab 3.2.
- b. Menentukan populasi dan sampel penelitian yang telah dibahas pada subbab 3.3.
- c. Menentukan variabel yang digunakan dalam penelitian, yaitu karakter, kapasitas, kapital, kekayaan, kondisi, dan kepatuhan seperti pada subbab 3.4.
- d. Merancang instrumen penelitian yang telah dibahas pada subbab 3.5.
- e. Melakukan uji coba instrumen penelitian.
- f. Pemeriksaan validitas dan reliabilitas instrumen penelitian yang telah dibahas pada subbab 2.9.
- g. Menyebar instrumen penelitian (kuesioner) yang telah valid dan reliabel.
- h. Melakukan metode pengukuran variabel yang telah dibahas pada subbab 2.8.
- i. Melakukan analisis *cluster* dengan langkah-langkah sebagai berikut:
 1. Melakukan pengujian multikolinearitas seperti dijelaskan pada sub subbab 2.4.1 menggunakan bantuan *software R 4.1.1*.



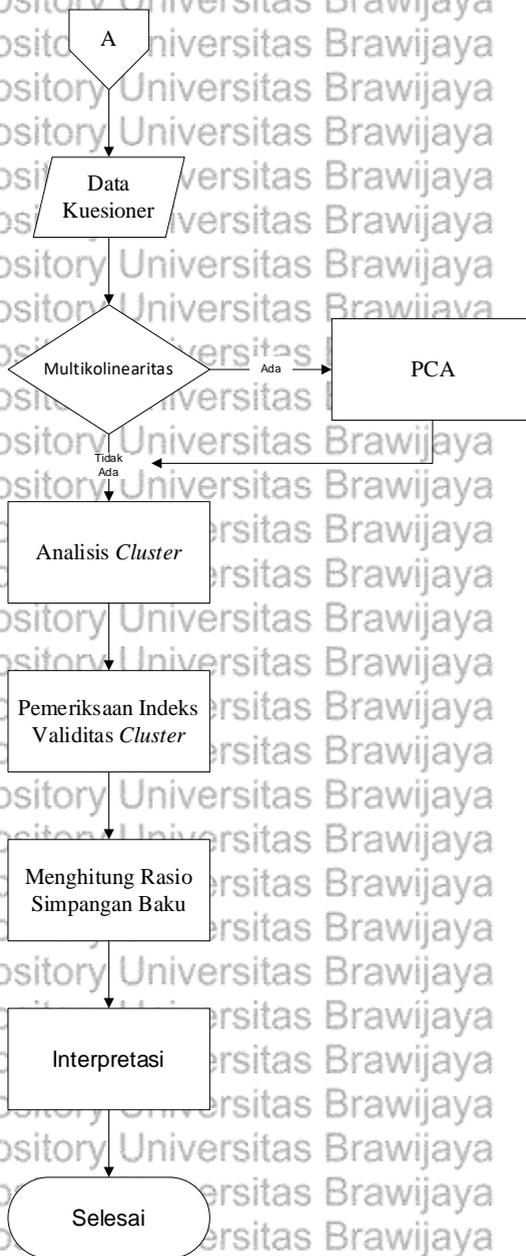
2. Jika terdapat korelasi antar variabel penelitian maka dilakukan Analisis Komponen Utama seperti pada sub subbab 2.4.2.
3. Melakukan pengukuran kesamaan menggunakan tiga metode pengukuran jarak secara bergantian seperti yang sudah dijelaskan pada subbab 2.2.
4. Melakukan pembentukan *cluster* menggunakan tiga metode pengelompokan terhadap ketiga hasil pengukuran jarak seperti yang telah dijelaskan pada subbab 2.3, sehingga akan menghasilkan 9 kombinasi metode pengukuran jarak dan metode pengelompokan.
5. Menghitung indeks validitas *cluster* seperti pada subbab 2.5 dengan *stopping rule* yang telah ditentukan.
6. Membandingkan kebaikan seluruh metode pengukuran jarak dan ketiga metode pengelompokan seperti pada subbab 2.6.
7. Interpretasi hasil.



3.8. Diagram Alir Penelitian



Gambar 3. 1. Diagram Alir Penelitian



Gambar 3.1. Lanjutan

BAB IV HASIL DAN PEMBAHASAN

4.1. Uji Multikolinieritas

Uji multiokolinieritas dilakukan untuk menguji apakah ditemukan adanya korelasi antar variabel. Pengujian ini perlu dilakukan karena pengukuran jarak menggunakan *Euclidean distance* dan *Manhattan distance* mengharuskan data terbebas dari korelasi antar variabel yang dapat menyebabkan *cluster* yang terbentuk akan tidak valid. Hasil uji multikolinieritas menggunakan *Variance Inflation Factor* (VIF) tersaji pada Tabel 4.1.

Tabel 4. 1. Hasil VIF dengan Y_1

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y_1		1,477	1,204	1,276	1,430	1,411
Y_2	1,331		1,153	1,282	1,390	1,417
Y_3	1,417	1,407		1,334	1,518	1,423
Y_4	1,279	1,426	1,216		1,507	1,415
Y_5	1,258	1,357	1,215	1,323		1,399
Y_6	1,296	1,446	1,190	1,298	1,462	

Jika nilai VIF ≥ 10 maka variabel terindikasi memiliki korelasi sehingga variabel tersebut harus dihilangkan. Tabel 4.1 menunjukkan bahwa seluruh nilai VIF < 10 , berarti tidak terjadi korelasi antar variabel pada data nasabah KPR Bank X. Sehingga tidak perlu melakukan analisis komponen utama pada data, dengan ini *Euclidean distance* dan *Manhattan distance* dapat digunakan untuk analisis *cluster* dan akan menghasilkan *cluster* yang valid.

4.2. Hasil Analisis Cluster

Analisis *cluster* bertujuan untuk mengelompokan dan sekumpulan objek yang beragam ke dalam kelompok (*cluster*) yang relatif homogen didasarkan pada karakteristik tertentu. Pada penelitian ini dilakukan analisis *cluster* hierarki *agglomerative* untuk mengelompokan nasabah KPR Bank X berdasarkan penilaian 5K menggunakan metode pengukuran jarak *Euclidean distance*, *Manhattan distance*, dan *Mahalanobis distance* dengan metode pengelompokan yaitu, *Ward method*, *Complete Linkage method*, dan *Average Linkage method* dengan bantuan *Software R 4.1.1*.



Penelitian ini akan menggunakan *stopping rule* sebesar 3 hingga 5 *cluster*. *Stopping rule* adalah batasan banyaknya *cluster* yang dijadikan pertimbangan peneliti sebelum menentukan banyaknya *cluster* paling optimal. Pada penelitian ini, penggunaan *stopping rule* yang terlalu kecil, yaitu 2 *cluster* akan menghasilkan *cluster* yang tidak baik, karena *cluster* 1 hanya berisi 1 hingga 10 anggota dan sisanya adalah anggota *cluster* 2. Sedangkan *stopping rule* yang terlalu besar, yaitu lebih dari 5 *cluster* akan menghasilkan *cluster* yang tidak valid, karena *cluster* yang terbentuk akan memiliki nilai *varians* antar *cluster* yang semakin kecil.

Hasil penghitungan jarak dengan *Euclidean distance* terdapat pada Lampiran, untuk *Manhattan distance* terdapat pada Lampiran, sedangkan untuk jarak *Mahalanobis* dapat dilihat pada Lampiran. Tahap selanjutnya dalam analisis *cluster* adalah memilih jumlah *cluster* yang optimal dari beberapa jumlah *cluster* yang telah dihasilkan. Pemilihan banyaknya *cluster* yang paling optimal akan menggunakan tiga indeks validitas *cluster*, yaitu indeks *Dunn*, *Davies Bouldin* dan *Silhouette*. Masing-masing indeks validitas memiliki prosedur dan ketentuan yang berbeda-beda dalam menentukan banyaknya *cluster* optimal, seperti yang telah dibahas pada subbab (2.5). Tahap terakhir adalah menentukan metode paling optimal untuk data nasabah KPR Bank X menggunakan rasio simpangan baku seperti pada subbab (2.6).

4.2.1. Hasil Analisis *Cluster* menggunakan *Euclidean distance* dengan metode *Ward*

Metode *Euclidean distance* akan mengukur jarak antar objek menggunakan persamaan (2.1), kemudian dikelompokkan menggunakan metode *Ward*, yaitu dengan mengambil nilai EES terkecil. Selanjutnya dilakukan penghitungan indeks validitas *cluster* dengan tiga hingga lima *cluster* sebagai *stopping rule*. Hasil indeks validitas *cluster* dari *Euclidean distance* dengan metode *Ward* tersaji pada Tabel 4.2.

Berdasarkan Tabel 4.2, dapat diketahui bahwa berdasarkan indeks *Dunn* dipilih empat *cluster* sebagai banyak *cluster* optimal karena memiliki nilai indeks *D* terbesar. Selanjutnya hasil yang berbeda ditunjukkan oleh indeks *Davies Bouldin*, terbentuknya tiga *cluster* merupakan banyak *cluster* yang optimal karena nilai *DB* yang

terkecil. Begitu pula dengan indeks *Silhouette*, tiga *cluster* adalah yang optimal karena nilai *S*-nya merupakan yang paling besar.

Tabel 4. 2. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Euclidean Distance* dengan metode *Ward*

Banyak <i>Cluster</i>	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,287	2,767	0,115
4	0,304	3,147	0,114
5	0,295	2,927	0,051

Setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* maka langkah selanjutnya yaitu menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Euclidean* dengan metode *Ward* dapat dilihat pada Tabel 4.3.

Tabel 4. 3. Hasil Peringkat Indeks Gabungan Jarak *Euclidean* dengan metode *Ward*

Banyak <i>Cluster</i>	Indeks <i>D</i>	<i>Rank</i>	Indeks <i>DB</i>	<i>Rank</i>	Indeks <i>S</i>	<i>Rank</i>	Indeks Gabungan
3	0,287	3	2,767	1	0,115	1	5
4	0,304	1	3,147	3	0,114	2	6
5	0,295	2	2,927	2	0,051	3	7

Berdasarkan Tabel 4.3, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) indeks *D*, indeks *DB*, dan indeks *S*. sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih tiga *cluster* sebagai banyak *cluster* optimal. Dendrogram hasil *Euclidean distance* dengan metode *Ward* sebanyak dua *cluster* tersaji pada Gambar 4.1.

Tabel 4. 5. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Euclidean Distance* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,289	2,672	0,090
4	0,300	2,581	0,080
5	0,315	3,018	0,073

Melihat pada Tabel 4.5, indeks *Dunn* dipilih lima cluster sebagai banyak cluster optimal karena memiliki nilai indeks *D* terbesar. Indeks *Davies Bouldin* menunjukkan bahwa banyak cluster optimal adalah sebanyak 4 cluster. Sedangkan pada indeks *Silhouette*, 3 cluster juga merupakan banyak cluster optimal

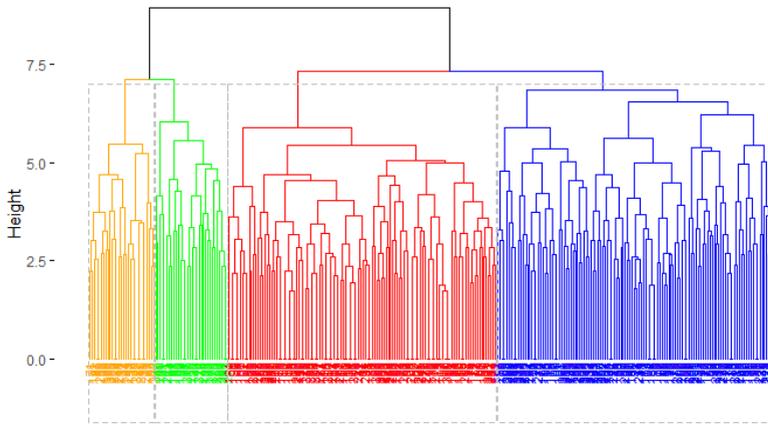
Masing-masing indeks validitas cluster telah menunjukkan banyaknya cluster optimal. Selanjutnya penentuan banyak cluster optimal ditentukan berdasarkan hasil penghitungan peringkat indeks gabungan. Hasil peringkat indeks gabungan dapat dilihat pada Tabel 4.6.

Tabel 4. 6. Hasil Peringkat Indeks Gabungan Jarak *Euclidean* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,289	3	2,672	3	0,090	1	7
4	0,300	2	2,581	1	0,080	2	5
5	0,315	1	3,018	2	0,073	3	6

Pada Tabel 4.6, clustering dengan *Euclidean distance* dengan *Complete Linkage* optimal harus memiliki empat cluster karena nilai indeks gabungannya terkecil yang diperoleh dari penjumlahan peringkat setiap indeks. Dengogram dari *Euclidean distance* dengan *Complete Linkage* sebanyak empat cluster dapat dilihat pada Gambar 4.2.

Cluster Dendrogram



Gambar 4. 2. Dendrogram *Euclidean Distance* dengan *Complete Linkage*

Gambar 4.2 menunjukkan dendrogram yang terbentuk dari penggunaan *Euclidean distance* dengan *Complete Linkage* sebanyak 4 cluster. Banyaknya anggota dari masing-masing cluster dapat dilihat pada Tabel 4.7. Anggota masing-masing cluster secara lengkap dapat dilihat pada Lampiran.

Tabel 4. 7. Banyak Anggota dari Empat Cluster Menggunakan *Euclidean Distance* dengan *Complete Linkage*

Cluster	Banyak Anggota
1	118
2	121
3	32
4	29

4.2.3. Hasil Analisis Cluster Menggunakan *Euclidean Distance* dengan *Average Linkage*

Metode *Euclidean distance* akan mengukur jarak antar objek menggunakan persamaan (2.1), kemudian dikelompokkan menggunakan metode *Average Linkage*, yaitu dengan menggunakan

rata-rata jarak. Hasil indeks validitas dari *Euclidean distance* dengan *Average Linkage* dapat dilihat pada Tabel 4.8.

Tabel 4. 8. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Euclidean Distance* dengan *Average Linkage*

Banyak Cluster	Indeks Dunn	Indeks Davies Bouldin	Indeks Silhouette
3	0,332	2,325	0,148
4	0,332	2,259	0,137
5	0,332	2,124	0,107

Tabel 4.8 menunjukkan bahwa nilai indeks *Dunn* pada setiap banyak *cluster* seluruhnya sama, namun banyak *cluster* yang optimal adalah tiga *cluster*. Selanjutnya hasil indeks *Davies Bouldin*, terbentuknya empat *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Pada indeks *Silhouette*, dua *cluster* adalah yang optimal karena nilai *S*-nya merupakan yang paling besar.

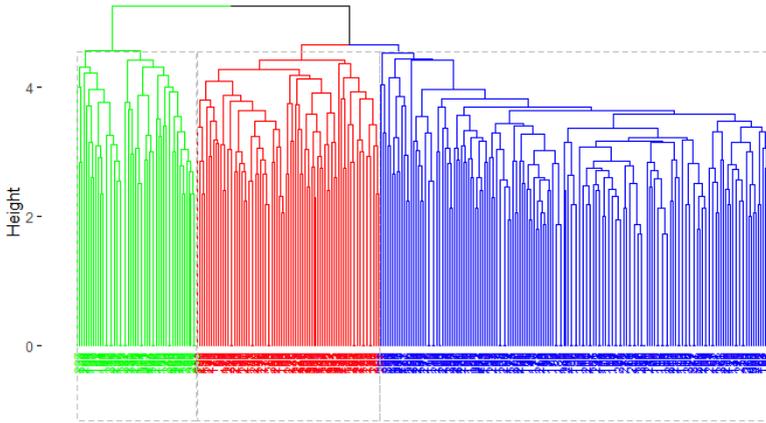
Langkah selanjutnya setelah mendapatkan banyak *cluster* optimal dari masing-masing indeks validitas *cluster* yang berbeda adalah menentukan banyaknya *cluster* optimal secara keseluruhan menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Euclidean* dengan *Average Linkage* dapat dilihat pada Tabel 4.9.

Tabel 4. 9. Hasil Peringkat Indeks Gabungan Jarak *Euclidean* dengan *Average Linkage*

Banyak Cluster	Indeks D	Rank	Indeks DB	Rank	Indeks S	Rank	Indeks Gabungan
3	0,332	1	2,325	3	0,148	1	5
4	0,332	2	2,259	2	0,137	2	6
5	0,332	2	2,124	1	0,107	3	6

Berdasarkan Tabel 4.9, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) setiap indeks. Dendrogram hasil *Euclidean distance* dengan *Average Linkage* sebanyak empat *cluster* tersaji pada Gambar 4.3.

Cluster Dendrogram



Gambar 4. 3. Dendrogram *Euclidean distance* dengan *Average Linkage*

Gambar 4.3 menunjukkan bentuk dendrogram menggunakan *Euclidean distance* dengan *Average Linkage* pada data kreditur KPR Bank X berdasarkan pada penilaian 6 K sebanyak 3 *cluster*. Banyaknya anggota dari masing-masing *cluster* dapat dilihat pada Tabel 4.10. Anggota masing-masing *cluster* dapat dilihat pada Lampiran.

Tabel 4. 10. Banyak Anggota dari Tiga *cluster* menggunakan *Euclidean distance* dengan *Average Linkage*

Cluster	Banyak Anggota
1	169
2	79
3	52

4.2.4. Hasil Analisis *Cluster* Menggunakan *Manhattan Distance* dengan metode *Ward*

Metode *Manhattan distance* akan mengukur jarak antar objek menggunakan persamaan (2.2), kemudian dikelompokkan menggunakan metode Metode *Ward*, yaitu dengan menggunakan nilai

SSE terkecil. Hasil indeks validitas dari *Manhattan distance* dengan Metode *Ward* dapat dilihat pada Tabel 4.11.

Tabel 4. 11. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Manhattan Distance* dengan Metode *Ward*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,291	2,793	0,168
4	0,255	3,330	0,081
5	0,255	2,991	0,063

Tabel 4.11 menunjukkan bahwa indeks *Dunn* pada tiga *cluster* memiliki nilai terbesar. Selanjutnya hasil yang sama ditunjukkan oleh indeks *Davies Bouldin*, terbentuknya tiga *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Sedangkan pada indeks *Silhouette*, tiga *cluster* adalah yang optimal karena nilai *S*-nya merupakan yang paling besar.

Setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* langkah selanjutnya yaitu menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Euclidean* dengan metode *Ward* dapat dilihat pada Tabel 4.12.

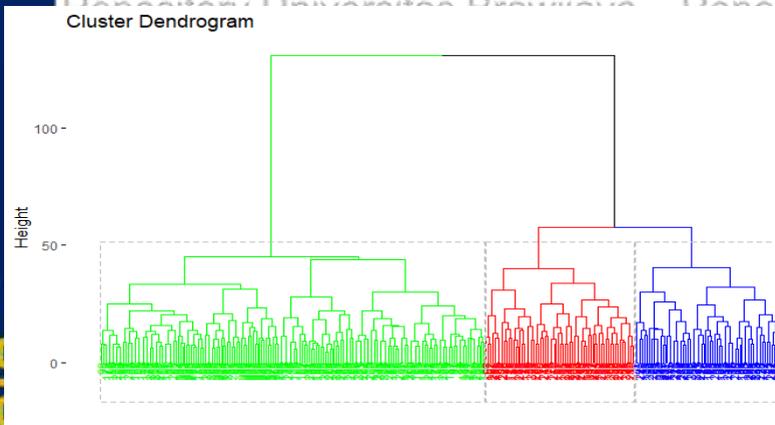
Tabel 4. 12. Hasil Peringkat Indeks Gabungan *Manhattan Distance* dengan Metode *Ward*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,291	1	2,793	1	0,168	1	3
4	0,255	2	3,330	3	0,081	2	7
5	0,255	2	2,991	2	0,063	3	7

Berdasarkan Tabel 4.12, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) indeks *D*, indeks *DB*, dan indeks *S*. sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih tiga *cluster* sebagai banyak



cluster optimal. Dendrogram hasil *Euclidean distance* dengan metode *Ward* sebanyak dua cluster tersaji pada Gambar 4.4.



Gambar 4. 4. Dendrogram *Manhattan distance* dengan metode *Ward*

Berdasarkan Gambar 4.4 dapat diketahui bahwa menggunakan *Manhattan distance* dengan metode *Ward* pada data kreditur KPR Bank X berdasarkan pada penilaian 6 K terbentuk 3 cluster. Banyaknya anggota dari masing-masing cluster dapat dilihat pada Tabel 4.13. Anggota masing-masing cluster dapat dilihat pada Lampiran.

Tabel 4. 13. Banyak Anggota dari Tiga cluster menggunakan *Manhattan Distance* dengan Metode *Ward*

Cluster	Banyak Anggota
1	170
2	66
3	64

4.2.5. Hasil Analisis Cluster Menggunakan *Manhattan Distance* dengan *Complete Linkage*

Metode *Manhattan distance* mengukur jarak antar objek menggunakan persamaan (2.2), kemudian dikelompokkan menggunakan metode *Complete Linkage*, yaitu menggunakan jarak terbesar. Hasil indeks validitas dari *Manhattan distance* dengan *Complete Linkage* dapat dilihat pada Tabel 4.14.

Tabel 4. 14. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Manhattan Distance* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,236	2,191	0,138
4	0,270	2,567	0,141
5	0,288	2,918	0,137

Melihat pada Tabel 4.14, indeks *Dunn* dipilih lima *cluster* sebagai banyak *cluster* optimal karena memiliki nilai indeks *D* terbesar. Pada indeks *Davies Bouldin* menunjukkan bahwa banyak *cluster* optimal adalah sebanyak 3 *cluster* yang ditunjukkan dengan nilai indeks terkecil. Sedangkan pada indeks *Silhouette*, 4 *cluster* merupakan banyak *cluster* optimal karena nilainya yang terbesar.

Masing-masing indeks validitas *cluter* telah menunjukkan banyaknya *cluster* optimal. Selanjutnya penentuan banyak *cluster* optimal ditentukan berdasarkan hasil penghitungan peringkat indeks gabungan. Hasil peringkat indeks gabungan dapat dilihat pada Tabel 4.15.

Tabel 4. 15. Hasil Peringkat Indeks Gabungan *Manhattan Distance* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,236	3	2,191	1	0,138	2	6
4	0,270	2	2,567	2	0,141	1	5
5	0,288	1	2,918	3	0,137	3	7

Pada Tabel 4.6, *clustering* dengan *Manhattan distance* dengan *Complete Linkage* optimal harus memiliki empat *cluster* karena nilai indeks gabungan terkecil yang diperoleh dari penjumlahan peringkat setiap indeks. Dengogram dari *Manhattan distance* dengan *Complete Linkage* sebanyak empat *cluster* dapat dilihat pada Gambar 4.5.

rata-rata jarak. Hasil indeks validitas dari *Manhattan distance* dengan *Average Linkage* dapat dilihat pada Tabel 4.17.

Tabel 4. 17. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Manhattan Distance* dengan *Average Linkage*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,254	1,557	0,172
4	0,291	1,888	0,158
5	0,291	1,841	0,137

Tabel 4.17 menunjukkan bahwa dari nilai indeks *Dunn* dipilih empat *cluster* sebagai banyak *cluster* optimal karena memiliki nilai indeks *D* terbesar. Hasil dari indeks *Davies Bouldin* adalah terbentuknya dua *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Selanjutnya dengan indeks *Silhouette*, tiga *cluster* adalah yang optimal karena nilai *S*-nya merupakan yang paling besar.

Langkah selanjutnya setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* yang berbeda adalah menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan *Manhattan distance* dengan *Average Linkage* dapat dilihat pada Tabel 4.18.

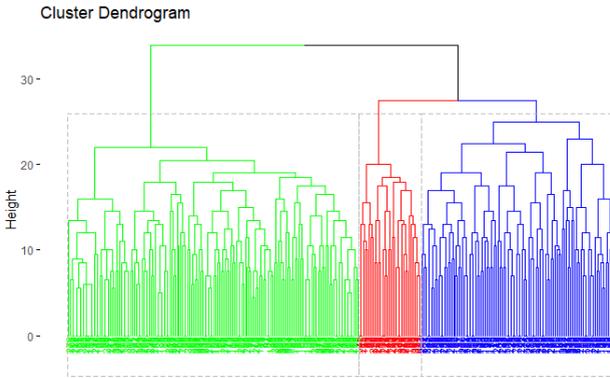
Tabel 4. 18. Hasil Peringkat Indeks Gabungan *Manhattan Distance* dengan *Average Linkage*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,254	3	1,557	1	0,172	1	5
4	0,291	1	1,888	3	0,158	2	6
5	0,291	2	1,841	2	0,137	3	7

Berdasarkan Tabel 4.18, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena setelah *rank* masing-masing *cluster* dijumlahkan, nilai indeks gabungan dari tiga *cluster* merupakan yang terkecil. Sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih tiga *cluster* sebagai banyak *cluster* optimal.



Dendrogram hasil *Manhattan distance* dengan *Average Linkage* sebanyak dua *cluster* tersaji pada Gambar 4.6.



Gambar 4. 6. Dendrogram *Manhattan Distance* dengan *Average Linkage*

Berdasarkan Gambar 4.6 dapat diketahui dendrogram yang terbentuk dari penggunaan *Manhattan distance* dengan *Average Linkage* pada data kreditur KPR Bank X berdasarkan pada penilaian 6 K sebanyak tiga *cluster*. Banyaknya anggota pada setiap *cluster* dapat dilihat pada tabel 4.19

Tabel 4. 19. Banyak Anggota pada 3 *Cluster* dengan *Manhattan Distance* dan *Average Linkage*

Cluster	Banyak Anggota
1	158
2	106
3	34

4.2.7. Hasil Analisis *Cluster* Menggunakan *Mahalanobis Distance* dengan Metode *Ward*

Metode *Mahalanobis distance* akan mengukur jarak antar objek menggunakan persamaan (2.3), kemudian dikelompokkan menggunakan metode metode *Ward*, yaitu dengan menggunakan nilai SSE terkecil. Hasil indeks validitas dari *Mahalanobis distance* dengan metode *Ward* dapat dilihat pada Tabel 4.20.

Tabel 4. 20. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Mahalanobis Distance* dengan metode *Ward*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,228	1,107	0,260
4	0,193	0,902	0,239
5	0,205	0,922	0,209

Berdasarkan indeks *Dunn* dipilih tiga *cluster* sebagai banyak *cluster* optimal, karena memiliki nilai Indeks *Dunn* terbesar. Selanjutnya hasil yang sama ditunjukkan oleh indeks *Davies Bouldin*, terbentuknya empat *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Begitu pula dengan indeks *Silhouette*, tiga *cluster* adalah banyak *cluster* optimal karena nilai *S*-nya merupakan yang paling besar.

Setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* yang berbeda maka langkah selanjutnya yaitu menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Mahalanobis* dengan metode *Ward* dapat dilihat pada Tabel 4.21.

Tabel 4. 21. Hasil Peringkat Indeks Gabungan *Mahalanobis Distance* dengan Metode *Ward*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,228	1	1,107	3	0,260	1	5
4	0,193	3	0,902	1	0,239	4	8
5	0,205	2	0,922	2	0,209	3	7

Berdasarkan Tabel 4.21, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) indeks *D*, indeks *DB*, dan indeks *S*. sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih tiga *cluster* sebagai banyak *cluster* optimal. Dendogram hasil *Mahalanobis distance* dengan metode *Ward* sebanyak dua *cluster* tersaji pada Gambar 4.7.

Tabel 4. 23. Indeks *Dunn*, *Davies Bouldin*, dan *Silhouette* pada *Mahalanobis Distance* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>Dunn</i>	Indeks <i>Davies Bouldin</i>	Indeks <i>Silhouette</i>
3	0,557	3,018	0,117
4	0,557	2,940	0,099
5	0,539	3,254	0,078

Tabel 4.23 menunjukkan bahwa berdasarkan indeks *Dunn* dipilih tiga *cluster* sebagai banyak *cluster* optimal.. Selanjutnya hasil yang ditunjukkan oleh indeks *Davies Bouldin*, terbentuknya empat *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Begitu pula dengan indeks *Silhouette*, tiga *cluster* adalah banyak optimal karena nilai *S*-nya merupakan yang terbesar.

Setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* yang berbeda maka langkah selanjutnya yaitu menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Mahalanobis* dengan *Complete Linkage* dapat dilihat pada Tabel 4.24.

Tabel 4. 24. Hasil Peringkat Indeks Gabungan Jarak *Mahalanobis* dengan *Complete Linkage*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,557	1	3,018	2	0,117	1	4
4	0,557	2	2,940	1	0,099	2	5
5	0,539	3	3,254	3	0,078	3	9

Berdasarkan Tabel 4.24, banyaknya *cluster* optimal adalah sebanyak tiga *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) indeks *D*, indeks *DB*, dan indeks *S*. sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih tiga *cluster* sebagai banyak *cluster* optimal. Dendogram hasil *Mahalanobis distance* dengan *Complete Linkage* sebanyak tiga *cluster* tersaji pada Gambar 4.8.

Banyak Cluster	Indeks Dunn	Indeks Davies Bouldin	Indeks Silhouette
3	0,282	2,795	0,168
4	0,255	3,310	0,081
5	0,234	2,931	0,063

Tabel 4.26 menunjukkan bahwa berdasarkan indeks *Dunn* dipilih tiga *cluster* sebagai banyak *cluster* optimal karena memiliki nilai indeks *D* terbesar. Selanjutnya hasil yang ditunjukkan oleh indeks *Davies Bouldin*, terbentuknya tiga *cluster* merupakan jumlah optimal karena nilai *DB* yang terkecil. Begitu pula dengan indeks *Silhouette*, tiga *cluster* adalah yang optimal karena nilai *S*-nya merupakan yang terbesar.

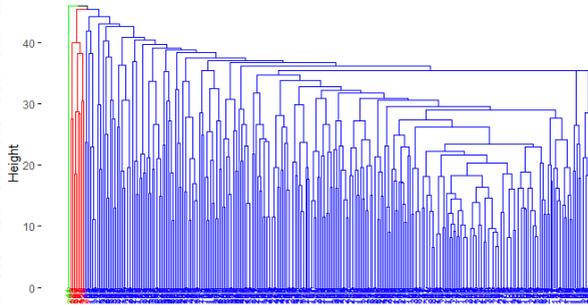
Setelah didapatkan keputusan dari masing-masing indeks validitas *cluster* yang berbeda maka langkah selanjutnya yaitu menentukan banyaknya *cluster* optimal menggunakan indeks gabungan, seperti pada subbab (2.5). Hasil penghitungan peringkat indeks gabungan untuk penggunaan jarak *Mahalanobis* dengan *Average Linkage* dapat dilihat pada Tabel 4.27.

Tabel 4. 27. Hasil Peringkat Indeks Gabungan Jarak *Mahalanobis* dengan *Average Linkage*

Banyak Cluster	Indeks <i>D</i>	Rank	Indeks <i>DB</i>	Rank	Indeks <i>S</i>	Rank	Indeks Gabungan
3	0,282	1	2,795	1	0,168	1	3
4	0,255	2	3,310	3	0,081	2	7
5	0,234	3	2,931	2	0,063	3	8

Berdasarkan Tabel 4.27, banyaknya *cluster* optimal adalah sebanyak dua *cluster* karena memiliki nilai indeks gabungan terkecil yang diperoleh dari hasil penjumlahan peringkat (*rank*) indeks *D*, indeks *DB*, dan indeks *S*. sehingga dapat diketahui bahwa berdasarkan indeks gabungan terpilih dua *cluster* sebagai banyak *cluster* optimal. Dendogram hasil *Mahalanobis distance* dengan *Complete Linkage* sebanyak dua *cluster* tersaji pada Gambar 4.9.

Cluster Dendrogram



Gambar 4. 9. Dendrogram Mahalanobis Distance dengan Average Linkage

Berdasarkan Gambar 4.9 dapat diketahui bahwa menggunakan Mahalanobis distance dengan metode Ward pada data kreditur KPR Bank X berdasarkan pada penilaian 6 Kterbentuk dua cluster. Jumlah anggota masing-masing cluster dapat dilihat pada tabel 4.28.

Tabel 4. 28. Banyak Anggota dari Tiga Cluster Menggunakan Mahalanobis Distance dengan Average Linkage

Cluster	Banyak Anggota
1	289
2	2
3	8

4.3. Pemilihan Hasil Clustering Terbaik

Pemilihan hasil clustering terbaik berguna untuk mengetahui seberapa baik kinerja seluruh metode yang digunakan untuk mengelompokkan objek penelitian. Suatu cluster atau kelompok dikatakan baik apabila memiliki kehomogenan tinggi dalam cluster dan keheterogenan yang tinggi antar cluster. Cara mengetahui kehomogenan dalam cluster yaitu menghitung simpangan baku dalam kelompok (S_w) seperti pada persamaan (2.11). Sedangkan untuk mengetahui keheterogenan antar cluster menggunakan simpangan baku antar kelompok (S_b) sesuai pada persamaan (2.12). Sehingga

hasil pengelompokan terbaik memiliki nilai rasio simpangan baku antar kelompok (S_b) dan simpangan baku dalam kelompok (S_w) yang maksimum, seperti disajikan pada persamaan (2.13). Hasil rasio simpangan baku disajikan pada Tabel 4.29.

Tabel 4. 29. Rasio Simpangan Baku Dalam *Cluster* dan Antar *Cluster*

Metode Pengukuran Jarak	Metode Pengelompokan	Simpangan Baku		Rasio Simpangan Baku
		S_w	S_b	
<i>Euclidean Distance</i>	Metode <i>Ward</i>	0,990	0,365	0,368
	Complete Linkage	0,850	0,400	0,471
<i>Euclidean Distance</i>	Average Linkage	0,968	0,324	0,329
<i>Manhattan Distance</i>	Metode <i>Ward</i>	0,967	0,377	0,390
	Complete Linkage	0,838	0,528	0,630
	Average Linkage	0,965	0,609	0,631
<i>Mahalanobis Distance</i>	Metode <i>Ward</i>	0,958	0,624	0,652
	Complete Linkage	0,830	0,526	0,634
	Average Linkage	0,974	0,406	0,417

Berdasarkan Tabel 4.29 diketahui bahwa metode yang menghasilkan nilai rasio simpangan baku antar kelompok (S_b) dan simpangan baku dalam kelompok (S_w) terbesar adalah metode *Mahalanobis distance* dengan metode *Ward* sebesar 0,652. Semakin besar nilai rasio simpangan baku dari suatu metode *cluster* maka semakin baik kinerja metode *cluster* dalam mengelompokkan objek. Sehingga kinerja metode *Mahalanobis distance* dengan metode *Ward* dapat dikatakan lebih baik dibandingkan dengan kinerja kombinasi metode-metode lain dalam pengukuran jarak dan pengelompokan nasabah KPR di Bank X tahun 2021 berdasarkan penilaian 6 K.



4.4. Interpretasi Hasil *Cluster* Menggunakan *Mahalanobis Distance* dengan metode *Ward*

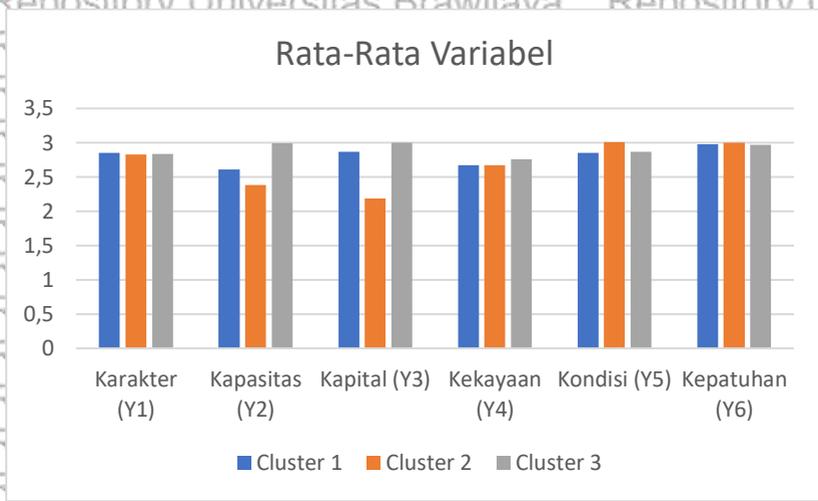
Pada subbab 4.3, *Mahalanobis distance* dengan metode *Ward* merupakan metode *cluster* hierarki yang paling optimal dengan *cluster* sebanyak 3 *cluster*. Masing-masing *cluster* memiliki karakteristik yang berbeda. Penghitungan nilai rata-rata masing-masing variabel pada setiap *cluster* seperti pada Tabel 4.30.

Tabel 4. 30. Rata-Rata Variabel pada Setiap *Cluster*

Variabel	Cluster 1	Cluster 2	Cluster 3
Karakter (Y ₁)	2,8513	2,8324	2,8365
Kapasitas (Y ₂)	2,6112	2,3841	2,9946
Kapital (Y ₃)	2,8647	2,1889	3,0033
Kekayaan (Y ₄)	2,6745	2,6710	2,7580
Kondisi (Y ₅)	2,8556	3,0097	2,8671
Kepatuhan (Y ₆)	2,9785	3,0043	2,9730
Rata-Rata	2,805	2,681	2,905

Karakteristik pada masing-masing *cluster* dapat ditunjukkan dengan menghitung rata-rata setiap variabel pada masing-masing *cluster*. Berdasarkan Tabel 4.23 dapat diketahui bahwa nilai rata-rata secara keseluruhan semua variabel penilaian 6 K sudah berada pada kondisi cukup (nilai rata-rata > 2,5). Nilai rata-rata setiap variabel dapat dilihat dalam bentuk diagram batang pada gambar 4.8.





Gambar 4. 8. Rata-Rata Variabel Penilaian 6 K dari Hasil *Clustering* menggunakan *Mahalanobis Distance* dengan *Ward's Method*

Berdasarkan Gambar 4.8 dapat diketahui bahwa *cluster* 1 memiliki penilaian karakter yang sedikit lebih tinggi, namun sedikit rendah pada kapasitas dan kekayaan dibandingkan *cluster* lain. Sehingga nasabah pada *cluster* 1 dapat dianggap memiliki karakter atau kepribadian yang paling baik jika dibandingkan dengan nasabah yang berada di *cluster* lain, namun nilai kapasitas atau kemampuan nasabah untuk membayar angsuran dan aset kekayaan yang nasabah miliki lebih rendah daripada nasabah di *cluster* 3 tetapi lebih tinggi jika dibandingkan dengan nasabah di *cluster* 2.

Nasabah yang berada pada *cluster* dua memiliki tingkat penilaian kondisi dan kepatuhan tertinggi diantara *cluster* yang ada. Namun tingkat karakter, kapasitas, kapital, dan kekayaan berada pada tingkat yang paling rendah. Artinya nasabah pada *cluster* dua memiliki kondisi ekonomi dan kepatuhan waktu membayar angsuran yang baik. Penilaian kepribadian dan aset kekayaan yang dimiliki nasabah di *cluster* 2 sedikit lebih rendah, sedangkan kemampuan nasabah dalam membayar angsuran dan kondisi eksternal yang dimiliki nasabah di *cluster* 2 merupakan yang terendah dengan selisih yang cukup jauh



dibandingkan dengan nasabah di *cluster* lain. *Cluster* 2 dapat dianggap sebagai *cluster* dengan anggota dengan penilaian 6 K yang paling rendah dikarenakan *cluster* 2 memiliki nilai terendah pada empat dari enam variabel yang digunakan.

Pada *cluster* tiga, nilai kapasitas, kapital, dan kekayaan berada pada tingkat yang paling tinggi, sedangkan pada penilaian yang lain sedikit lebih rendah dibandingkan *cluster* lainnya. Kesimpulan pada *cluster* tiga adalah nasabah yang merupakan anggota *cluster* tiga memiliki tingkat kemampuan membayar angsuran, kondisi keuangan, dan kondisi eksternal yang tertinggi jika dibandingkan dengan anggota di *cluster* yang lain. *Cluster* tiga memiliki penilaian kepribadian, kondisi eksternal, dan kepatuhan waktu membayar angsuran yang sedikit lebih rendah tetapi bukan yang terendah diantara ketiga *cluster* yang terbentuk. Sehingga *cluster* tiga dapat dianggap sebagai *cluster* dengan anggota yang memiliki penilaian 6 K paling baik dibandingkan anggota pada *cluster* lain.

BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan

Pada penelitian ini kesimpulan yang dapat diambil adalah sebagai berikut:

1. Data nasabah KPR di Bank X berdasarkan penilaian 6 K menggunakan *Euclidean distance – Ward's method, Euclidean distance – Average Linkage method, Manhattan distance – Ward's method, Manhattan distance – Average Linkage method, Mahalanobis distance – Ward's method, Mahalanobis distance – Complete Linkage method, dan Mahalanobis distance – Average Linkage method* terbentuk *cluster* optimal sebanyak 3 *cluster*; sedangkan dengan *Euclidean distance – Complete Linkage method dan Manhattan distance – Complete Linkage method* menghasilkan 4 *cluster* optimal.
2. Kinerja metode *clustering* yang paling baik untuk mengelompokan nasabah KPR di Bank X tahun 2021 adalah metode *Mahalanobis distance* dengan metode *Ward*.

5.2. Saran

Beberapa saran yang dapat diberikan berdasarkan hasil penelitian ini adalah sebagai berikut:

1. Pihak Bank X dapat memeriksa kembali kemampuan nasabah untuk melunasi angsuran KPR berdasarkan *cluster* yang sudah terbentuk.
2. Pihak Bank X dapat mengevaluasi penerimaan calon nasabah KPR di Bank X.
3. Peneliti selanjutnya dapat menggunakan metode pengukuran *Mahalanobis distance* dengan metode metode *Ward* karena hasil penelitian menunjukkan kombinasi tersebut merupakan yang paling baik diantara lainnya.
4. Penelitian selanjutnya dapat menggunakan metode pengukuran selain *Euclidean distance, Manhattan distance, dan Mahalanobis distance*.
5. Penelitian selanjutnya dapat menggunakan indeks validitas *cluster* selain indeks *Dunn, Davies Bouldin, dan Silhouette*.

DAFTAR PUSTAKA

Astiko dan Sunardi. 1996. *Pengantar Manajemen Perkreditan*. Edisi pertama. Yogyakarta: ANDI.

Astutik, S., Solimun dan Darmanto. 2018. *Analisis Multivariat, Teori dan Aplikasinya dengan SAS*. Malang: UB Press

Azuaje, F dan Bolshakova, N. 2001. *Improving Expression Data Mining Through Cluster Validity*. Ireland: Dublin Publication Inc.

Budiarto, Eko. 2004. *Metodologi Penelitian Kedokteran: Sebuah Pengantar*. Jakarta: EGC.

Dubes and Jain, A. K. 1988. *Algorithm for Clustering Data*. New Jersey: Prentice Hall.

Ghozali, I. 2018. *Aplikasi Analisis Multivariat SPSS 25*. Ninth edition. Semarang: Universitas Diponegoro.

Hair, Joseph F., Anderson, Rolph E. Tatham, R.L. dan Black, William C. 2006. *Multivariate Data Analysis*. Sixth edition. New Jersey: Pearson Education Inc.

Hair, Joseph F., Black, William C., Babin, Barry J. dan Anderson, Rolph E. 2010. *Multivariate Data Analysis*. Seventh edition. Boston: Pearson.

Hardle, Wolfgang K., dan Simar, Leopold. 2003. *Applied Multivariate Statistical Analysis*. Berlin: Springe-Verlag.

Hardle, Wolfgang K., dan Simar, Leopold. 2012. *Applied Multivariate Statistical Analysis*. Third edition. Berlin: Springer.

Johnson, R. A. dan Wichern D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth edition. Boston: Pearson Education, Inc.



Kline, P. 2000. *The Handbook of Psychological Testing*. Second edition. London: Routledge, New Fetter Lane.

Kovacs, F., Legany, C., dan Babos, A. 2005. "Cluster Validity Measurement Techniques". *Jurnal 6th International Smposium of Hungarian Researchers on Computational Intelligence*. Budapest, Hungary.

Mattjik, A., dkk. 2002. *Aplikasi Analisis Peubah Ganda*. Bogor: Depdiknas.

Priyatno, Duwi. 2012. *Cara Kilat Belajar Analisis Data dengan SPSS 20*. Yogyakarta: Andi Offset.

Rencher, Alvin C. 2002. *Methods of Multivariate Analysis*. Second Edition. New York: John Wiley & Sons, Inc.

Riduwan M. B. A. 2005. *Skala Pengukuran Variabel-Variabel Penelitian*. Cetakan ketiga. Bandung: CVc Alfabeta.

Seber, G. A. F. 2004. *Multivariate Obseervations*. New Jersey: John Wiley & Sons, Inc.

Su, M.C. 2003. "A New Index of Cluster Validit". *Jurnal Electrical Engineering*. Tamkang, Taiwan.

Sudjana. 2004. *Metoda Statistika*. Bandung: PT. Tarsito.

Sugiyono. 2018. *Metode Penelitian Kuantitatif*. Bandung: CV. Alfabeta.

Supranto. Honases. 2004. *Analisis ultivariat Arti & Interpretasi*. Jakarta.: PT. Rineka Cipta.

Suwarno, Bambang. 2005. *Pengantar Aplikasi Statistika*. Bandung: PPs Universitas Pendidikan Indonesia.

Solimun. 2010. *Analisis Multivariat Pemodelan Struktural Metode Partial Least Square -PLS*. Malang: CV. Citra Malang.



Solimun., Fernandes, A. A. R. dan Nurjannah. 2017. *Metode Statistika Multivariat. Pemodelan Persamaan Struktural (SEM) Pendekatan WarpPLS*. Malang: UB Press.

Walpole, Ronald E., Myers, Ramond H., Myers, Sharon L. 2012. *Probability & Statistics for Engineers & Scientists*. Ninth edition. Boston: Pearson Education, Inc.

Yamin, S.R. 2011. *Regresi dan Korelasi dalam Genggaman Anda*. Jakarta: Salemba Empat.





LAMPIRAN

Lampiran 1. Kuesioner Penelitian

VARIABEL KARAKTER						
No.	Pernyataan	Skala Sikap				
		Sangat tidak setuju ←...→ Sangat Setuju				
Itikad dan Tanggung Jawab		1 STS	2 TS	3 N	4 S	5 SS
1.	Saya secara rutin membayar cicilan Kredit Pemilikan Rumah tepat waktu					
2.	Saya berusaha agar tidak pernah telat dalam membayar cicilan Kredit Pemilikan Rumah					
Sifat atau Watak						
1.	Saya menghabiskan sebagian besar penghasilan untuk keperluan keluarga					
2.	Saya menabung kelebihan penghasilan					
Komitmen Pembayaran						
1.	Saya merasa bersalah jika tidak membayar cicilan Kredit Pemilikan Rumah tepat waktu					
2.	Saya merasa harus membayar cicilan tepat waktu karena pengajuan kredit disetujui					

VARIABEL KAPASITAS

Pendapatan Nasabah	STS	TS	N	S	SS
1. Saya memiliki penghasilan yang kurang dari standar UMR					
2. Pendapatan saya tidak hanya berasal dari satu sumber					

Kemampuan dalam Membayar Angsuran

1. Saya mampu membayar cicilan secara berkala					
2. Saya harus menjual/berhutang untuk membayar angsuran kredit					

Kemampuan dalam Menyelesaikan Kredit Tepat Waktu

1. Saya siap menerima sanksi jika tidak mampu membayar cicilan Kredit Pemilikan Rumah					
2. Saya mampu untuk melakukan pembayaran cicilan tepat waktu					

VARIABEL KAPITAL (MODAL)

Sumber Penghasilan Tetap	STS	TS	N	S	SS
1. Saya memiliki penghasilan yang tetap					
2. Saya memiliki pekerjaan yang tetap					



miliki Bidang Usaha Lain sebagai Pneghasilan Sumber		STS	TS	N	S	SS
1.	Saya memiliki pekerjaan lain untuk membayar cicilan Kredit Pemilikan Rumah					
2.	Saya memiliki usaha sendiri untuk meningkatkan penghasilan					
Memiliki Tabungan atau Simpanan di Bank						
1.	Saya selalu memiliki cadangan tabungan agar dapat membayar cicilan Kredit Pemilikan Rumah tepat waktu					
2.	Saya memiliki cadangan tabungan barang agar dapat membayar cicilan Kredit Pemilikan Rumah tepat waktu					
VARIABEL KEKAYAAN						
Nilai jual barang jaminan yang digunakan sebanding/melebihi plafond kredit						
1.	Saya memiliki barang jaminan yang berharga untuk dijadikan jaminan jika saya tidak dapat melunasi Kredit Pemilikan Rumah					



2.	Saya memiliki jaminan yang sebanding dengan nilai KPR					
Jaminan Bersifat Fisik atau Non-fisik		STS	TS	N	S	SS
1.	Saya masih memiliki jaminan fisik jika saya gagal melunasi Kredit Pemilikan Rumah					
2.	Jika saya gagal melunasi Kredit Pemilikan Rumah saya merelakan aset berharga saya untuk dilelang oleh pihak bank					
Status kepemilikan barang jaminan dan keaslian dokumen						
1.	Saya memiliki bukti keaslian aset berharga yang menjadi jaminan yaitu berupa sertifikat					
2.	Seluruh aset yang saya miliki merupakan milik pribadi					
VARIABEL KONDISI						
Pengembangan Bisnis/Usaha/Investasi						
1.	Usaha yang saya jalani memiliki prospek yang baik dalam jangka panjang					
2.	Mudah untuk mengembangkan usaha					



	yang akan saya jalankan				
Fluktuasi Perekonomian					
1.	Pekerjaan yang saya lakukan masih dapat bertahan meskipun ada ancaman krisis ekonomi				
2.	Sulit dilakukan pengembangan usaha yang sedang saya jalankan				
Kondisi Sosial Ekonomi					
1.	Keluarga saya termasuk keluarga yang berkecukupan, sehingga penghasilan mampu untuk membayar cicilan Kredit Pemilikan Rumah tepat waktu				
2.	Pembayaran untuk pendidikan anak saya termasuk murah, sehingga saya tidak kesulitan untuk membayar cicilan Kredit Pemilikan Rumah tepat waktu				
VARIABEL KEPATUHAN					
Ketepatan Waktu					
1.	Saya selalu ingin membayar cicilan Kredit Pemilikan Rumah tepat pada waktunya				



2.	Saya selalu membayar cicilan Kredit Pemilikan Rumah sebelum jatuh tempo				
Akurasi Data					
1.	Saya selalu membayar cicilan Kredit Pemilikan Rumah sesuai dengan besaran cicilan				
2.	Saya selalu melunasi kredit yang saya ambil				
Sanksi					
1.	Sanksi dibuat untuk menumbuhkan kedisiplinan dari nasabah				
2.	Sanksi digunakan sebagai efek jera bagi yang melanggar				

**Lampiran 2. Data Variabel**

No	ID Bank	X111	X112	X121	...	Y131	Y132
1	10	4	5	3	...	4	4
2	10	5	2	3	...	3	5
3	10	3	1	4	...	2	3
4	10	5	3	3	...	5	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
101	30	2	4	2	...	3	4
102	30	4	1	2	...	2	3
103	30	4	4	4	...	4	5
104	30	4	4	4	...	2	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
297	43	4	4	3	⋮	3	2
298	43	4	4	3	...	3	4
299	43	5	4	3	...	4	2
300	43	5	3	4	...	4	4



Lampiran 3. Coding Analisis *cluster* dan Indeks Validitaas dengan Software R

```
library(readxl)
library(cIValid)
library(biotools)
library(cluster)
library(dendextend)
library(gtable)
library(NbClust)
library(MASS)
library(factoextra)
library(dplyr)

datax <- read_excel("D:/Data/300baru.xlsx")
datax %>% head()
data=as.data.frame(datax)
rownames(data) <- c(1:300)
data %>% head()

## indeks Dunn & Silhouette
dun_sil1 = cIValid (data, 3:5, cIMethods = "hierarchical", validation
="internal", metric = "euclidean", method = "ward.D2")
summary(dun_sil1)
dun_sil2 = cIValid (data, 3:5, cIMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = " ward.D2")
summary(dun_sil2)
dun_sil3 = cIValid (data, 3:5, cIMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = " ward.D2")
summary(dun_sil3)
dun_sil4 = cIValid (data, 3:5, cIMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "complete")
```





```
summary(dun_sil4)
dun_sil5 = clValid (data, 3:5, clMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "complete")
summary(dun_sil5)
dun_sil6 = clValid (data, 3:5, clMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "complete")
summary(dun_sil6)
dun_sil7 = clValid (data, 3:5, clMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "average")
summary(dun_sil7)
dun_sil8 = clValid (data, 3:5, clMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "average")
summary(dun_sil8)
dun_sil9 = clValid (data, 3:5, clMethods = "hierarchical", validation
="internal", metric = " euclidean ", method = "average")
summary(dun_sil9)

## indeks Davies Bouldin & cluster
clus_hier11 = eclust (data, FUNcluster = "hclust", k = 3, hc_metric =
"euclidean", hc_method = "ward.D2", graph = TRUE, scale =
"none")
cl1 = clus_hier11$cluster
dav_index1 = index.DB (data ,cl1 ,d = NULL ,centrotypes =
"centroids", p=2 ,q=2)
dav_index1$DB
dend1 = fviz_dend (clus_hier11 ,horiz = FALSE ,rect=TRUE ,cex=0.6
, k=3, k_color=c("green","red","blue"))
dend1
x<-as.matrix(data)
stopifnot(mahalanobis(x, 0, diag(ncol(x))) == rowSums(x*x))
center <- colMeans(x)
cov <- cov(x)
```



cov

```
mahadis <- D2.dist(x,cov)
```

```
dendogram1 = hclust(mahadis,"ward.D2")
```

```
dendogram1
```

```
print(plot(dendogram3), hang=-1, cex=0.6)
```

```
dend1= fviz_dend(dendogram1, rect=TRUE, hang=-1, cex=0.6, k=3,
```

```
k_color=c("green","red","blue"))
```

```
dend1
```





Lampiran 4. Coding Keباikan Metode Cluster

```
## rasio simpangan baku
euward=cbind(data,cl1)
euward1= euward [which(euward$cl5=="1"),]
euward2= euward [which(euward$cl5=="2"),]
euward3= euward [which(euward$cl5=="3"),]
std_w_1=sd(as.matrix(euward1[,-20]))
std_w_2=sd(as.matrix(euward2[,-20]))
std_w_3=sd(as.matrix(euward3[,-20]))
m_w_1=mean(as.matrix(euward1[,-20]))
m_w_2=mean(as.matrix(euward2[,-20]))
m_w_3=mean(as.matrix(euward3[,-20]))
m_all=mean(m_w_1,m_w_2,m_w_3)
sw=(std_w_1+std_w_2+std_w_3)/3
sb=sqrt(((m_w_1-m_all)^2+(m_w_2-m_all)^2+(m_w_3-
m_all)^2)/3)
sb
rasio_ward=sw/sb
rasio_ward

## statsitika deskriptif
rata_1=apply(ward_1[,-20],2,mean)
rata_2=apply(ward_2[,-20],2,mean)
rata_3=apply(ward_3[,-20],2,mean)
rata=rbind(rata_1,rata_2,rata_3)
rata
```

Lampiran 5. Output Indeks Validitas *Dunn* dan *Sillhouette*

a. *Euclidean distance* dengan *Ward's Method*

```

Clustering Methods:
  hierarchical

Cluster sizes:
  3 4 5

Validation Measures:
                                     3         4         5

hierarchical connectivity 159.1440 169.8393 221.2143
      Dunn                0.2873  0.3048  0.2957
      Silhouette          0.1157  0.1144  0.0515
    
```

	Score <dbl>	Method <chr>	Clusters <chr>
Connectivity	159.1440	hierarchical	3
Dunn	0.3048	hierarchical	4
Silhouette	0.1157	hierarchical	3

b. *Euclidean distance* dengan *Complete Linkage*

```

Clustering Methods:
  hierarchical

Cluster sizes:
  3 4 5

Validation Measures:
                                     3         4         5

hierarchical connectivity 185.7377 204.0762 234.8778
      Dunn                0.2894  0.3007  0.3153
      Silhouette          0.0909  0.0802  0.0736
    
```

	Score <dbl>	Method <chr>	Clusters <chr>
Connectivity	185.7377	hierarchical	3
Dunn	0.3153	hierarchical	5
Silhouette	0.0909	hierarchical	3

c. *Euclidean distance dengan Average Linkage*Clustering Methods:
hierarchicalCluster sizes:
3 4 5

Validation Measures:

		3	4	5
hierarchical	Connectivity	126.3175	136.4167	142.2607
	Dunn	0.3324	0.3324	0.3324
	Silhouette	0.1481	0.1378	0.1078

	Score <dbl>	Method <chr>	Clusters <chr>
Connectivity	126.3175	hierarchical	3
Dunn	0.3324	hierarchical	3
Silhouette	0.1481	hierarchical	3