

**ANALISIS *CLUSTER K-MODES* DENGAN VALIDASI  
*DAVIES BOUILDIN INDEX* DALAM MENGELOMPOKKAN  
PROVINSI DI INDONESIA BERDASARKAN INDIKATOR  
RUMAH TANGGA MISKIN**

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar  
Sarjana Statistika*



Oleh  
**SYIFA AZAHRA**  
**NIM. 20337011**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2024**

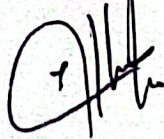
**PERSETUJUAN SKRIPSI**

**ANALISIS *CLUSTER K-MODES* DENGAN  
VALIDASI *DAVIES BOUILDIN INDEX* DALAM  
MENGELOMPOKKAN PROVINSI DI INDONESIA  
BERDASARKAN INDIKATOR RUMAH TANGGA  
MISKIN**

Nama : Syifa Azahra  
NIM : 20337011  
Program Studi : SI Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

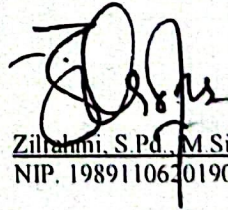
Padang, 30 Mei 2024

Mengetahui:  
Kepala Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si.  
NIP. 198402232010122005

Disetujui Oleh:  
Pembimbing



Zilrahmi, S.Pd., M.Si.  
NIP. 198911061019032009



## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Syifa Azahra  
NIM : 20337011  
Program Studi : SI Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### **ANALISIS *CLUSTER K-MODES* DENGAN VALIDASI *DAVIES BOUILDIN INDEX* DALAM MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR RUMAH TANGGA MISKIN**

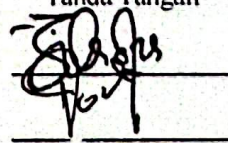
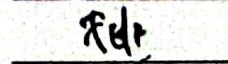

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 30 Mei 2024

#### Tim Penguji

	Nama
Ketua	: Zilrahmi, S.Pd., M.Si.
Anggota	: Dodi Vionanda, M.Si., Ph.D.
Anggota	: Fadhilah Fitri, S.Si., M.Stat.

Tanda Tangan

  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_

## **ABSTRAK**

### **Analisis Cluster K-modes dengan Validasi Davies Bouildin Index dalam Mengelompokan Provinsi Di Indonesia Berdasarkan Indikator Rumah Tangga Miskin**

**Oleh : Syifa Azahra**

Kemiskinan merupakan salah satu masalah sosial yang terjadi di Indonesia. Persentase penurunan kemiskinan di Indonesia pada tahun 1999 sampai tahun 2004 sekitar 6,77% dalam 5 tahun, namun dari tahun 2014 sampai tahun 2019 persentase penurunan kemiskinan hanya 1,74% saja. Angka ini menunjukkan terjadinya hambatan dari program penganggulangan kemiskinan di Indonesia. Maka dari itu penelitian ini membahas tentang pengelompokan provinsi di Indonesia berdasarkan indikator rumah tangga miskin. Tujuan penelitian ini untuk mengetahui provinsi mana saja yang masuk ke dalam cluster sama, untuk mengetahui karakteristik masing-masing cluster, dan untuk mengukur seberapa baik hasil algoritma dalam memisahkan amatan ke dalam kelompok-kelompok yang berbeda menggunakan jumlah cluster optimumnya. Penelitian ini merupakan penelitian terapan, yaitu penerapan metode K-Modes dengan validasi Davies Bouildin Index (DBI) dalam mengelompokan provinsi di Indonesia berdasarkan indikator rumah tangga miskin. Data yang digunakan merupakan data sekunder dari Survei Demografi dan Kesehatan Indonesia 2017 tentang Daftar Rumah Tangga. Variabel yang digunakan terdiri dari 12 variabel yang bertipe kategorik. Hasil penelitian menunjukkan terdapat 2 cluster, dimana jumlah cluster tersebut sudah tepat digunakan dalam algoritma dikarenakan memiliki nilai DBI sebesar 1,89. Dari 2 cluster tersebut, cluster 1 merupakan cluster yang memenuhi 9 indikator rumah tangga miskin dengan 8 provinsi yaitu Papua, Papua Barat, Nusa Tenggara Timur, Maluku, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat. Maka dari itu pemerintah perlu memprioritaskan 8 provinsi tersebut untuk menangani masalah kemiskinan di Indonesia berdasarkan indikator rumah tangga miskin yang ada pada cluster tersebut, agar target penurunan kemiskinan tercapai setiap tahunnya.

**Kata Kunci:** Davies Boudin Index, Kemiskinan, K-modes, Pengelompokan.

# **K-modes Analysis with Davies Bouildin Index Validation In Grouping Provinces in Indonesia Based on Indicator of Poor Households**

**Syifa Azahra**

## **ABSTRACT**

Poverty is one of the social problems that occur in Indonesia. The percentage of poverty reduction in Indonesia from 1999 to 2004 was around 6.77% in 5 years, but from 2014 to 2019 the percentage of poverty reduction was only 1.74%. This figure shows the obstacles of the poverty reduction program in Indonesia. Therefore, this research discusses the grouping of provinces in Indonesia based on indicators of poor households. The purpose of this research is to find out which provinces are included in the same cluster, to find out the characteristics of each cluster, and to measure how well the algorithm results in separating observations into different groups using the optimum number of clusters.

This research is an applied research, namely the application of the K-Modes method with validation of the Davies Bouildin Index (DBI) in clustering provinces in Indonesia based on indicators of poor households. The data used is secondary data from the 2017 Indonesian Demographic and Health Survey on Household Registration. The variables used consist of 12 variables of categorical type.

The results show that there are 2 clusters, where the number of clusters is appropriate for use in the algorithm because it has a DBI value of 1,89. Of the 2 clusters, cluster 1 is a cluster that fulfills 9 indicators of poor households with 8 provinces namely Papua, West Papua, East Nusa Tenggara, Maluku, North Maluku, Central Sulawesi, Southeast Sulawesi, West Sulawesi. Therefore, the government needs to prioritize these 8 provinces to deal with the problem of poverty in Indonesia based on the indicators of poor households in the cluster, so that the poverty reduction target is achieved every year.

**Kata Kunci:** Clustering, Davies Boudin Index, K-modes, Poverty

## KATA PENGANTAR

Alhamdulillah puji dan syukur penulis ucapkan kepada Allah SWT, yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul “Analisis *Cluster K-Modes* dengan Validasi *Davies Bouildin Index* dalam Mengelompokan Provinsi Di Indonesia Berdasarkan Indikator Rumah Tangga Miskin”.

Penulisan skripsi ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana pada program studi Sarjana Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang. Pada penulisan skripsi ini, penulis banyak mendapat dukungan, bimbingan, dan bantuan dari berbagai pihak. Dalam kesempatan ini, penulis mengucapkan terimakasih kepada:

1. Ibu Zilrahmi, S.Pd., M.Si., Dosen Pembimbing Skripsi dan Dosen Pembimbing Akademik yang telah memberikan bimbingan dan dorongan dalam menyelesaikan skripsi ini.
2. Bapak Dodi Vionanda, M.Si., Ph.D., Dosen Pembahas Skripsi dan Koordinator Prodi Statistika (DIII) yang telah memberikan arahan dan masukan dalam menyelesaikan skripsi.
3. Ibu Fadhilah Fitri S.Si., M.Stat., Dosen Pembahas Skripsi yang telah memberikan arahan dan masukan dalam menyelesaikan skripsi ini.
4. Bapak dan Ibu Dosen, Staf Pengajar Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang yang

telah membimbing dan berbagi ilmu pengetahuan kepada penulis selama duduk di bangku perkuliahan.

5. Teristimewa kepada kedua orang tua, adik dan keluarga besar yang selalu memberikan do'a dan semangat kepada penulis. Terimakasih sebesar-besarnya penulis ucapkan kepada beliau atas segala bantuan, motivasi, dan nasihat tiada henti hingga penulis dapat menyelesaikan studinya sampai sarjana.
6. Semua sahabat dan rekan-rekan seperjuangan yang telah banyak membantu dan memberi dukungan kepada penulis.

Semoga segala bimbingan, bantuan, dan motivasi yang telah diberikan menjadi amal kebaikan dan mendapatkan balasan dari Allah SWT. Penulis menyadari bahwa penulisan skripsi ini jauh dari kesempurnaan. Oleh karena itu, saran dan masukan dari berbagai pihak sangat diharapkan demi kesempurnaan penulisan yang akan datang. Akhir kata penulis berharap skripsi ini dapat bermanfaat bagi semua pihak.

Padang, 25 Mei 2024

Syifa Azahra

## DAFTAR ISI

ABSTRAK.....	i
KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	vi
DAFTAR TABEL.....	vii
DAFTAR LAMPIRAN.....	viii
BAB I PENDAHULUAN.....	1
A. Latar Belakang.....	1
B. Batasan Masalah.....	6
C. Rumusan Masalah.....	6
D. Tujuan Penelitian.....	7
E. Manfaat Penelitian.....	7
BAB II KERANGKA TEORITIS.....	8
A. Kajian Teori.....	8
B. Penelitian Terdahulu.....	23
BAB III METODOLOGI PENELITIAN.....	26
A. Jenis Penelitian.....	26
B. Jenis dan Sumber Data.....	26
C. Variabel Penelitian.....	26
D. Teknik Analisis Data.....	28
BAB IV HASIL DAN PEMBAHASAN.....	31
A. Hasil.....	31
B. Pembahasan.....	45
BAB V PENUTUP.....	48
A. Kesimpulan.....	48
B. Saran.....	49
DAFTAR PUSTAKA.....	50
LAMPIRAN.....	54



## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Perkembangan Kemiskinan di Indonesia.....	2
2. Tahapan <i>data mining</i> .....	13
3. Contoh hasil <i>clustering hierarki</i> .....	15
4. Contoh hasil <i>clustering non-hierarki</i> .....	16
5. Grafik metode <i>elbow</i> .....	17
6. Tahapan analisis.....	30
7. Eksplorasi data variabel motor.....	32
8. Ekplorasi data variabel mobil.....	32
9. Ekplorasi data variabel rekening bank.....	33
10. Ekplorasi data variabel listrik.....	33
11. Ekplorasi data variabel bahan lantai.....	34
12. Ekplorasi data variabel bahan dinding.....	34
13. Ekplorasi data variabel bahan atap.....	35
14. Ekplorasi data variabel sumber air minum.....	35
15. Ekplorasi data variabel jenis toilet.....	36
16. Ekplorasi data variabel anggota rumah tangga.....	36
17. Ekplorasi data variabel luas lantai.....	37
18. Ekplorasi data variabel bahan bakar memasak.....	37
19. Keterangan warna tiap provinsi.....	38
20. Jumlah <i>cluster</i> optimum.....	39

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Variabel penelitian.....	26
2. Sturktur data penelitian.....	28
3. <i>Centroid</i> awal.....	39
4. <i>Cluster</i> pertama.....	40
5. <i>Cluster</i> kedua.....	41
6. <i>Centroid</i> baru.....	41
7. Membandingkan <i>Centroid</i> .....	41
8. Hasil <i>clustering</i> .....	42
9. Karakteristik <i>cluster</i> .....	42
10. Nilai rasio.....	44

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
Lampiran 1. Hasil deteksi data <i>noise</i> .....	54
Lampiran 2. Hasil penanganan data <i>noise</i> .....	55
Lampiran 3. Data sudah agregat.....	56

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang**

Indonesia merupakan negara berkembang dengan tingkat pertumbuhan penduduk setiap tahunnya cukup tinggi dimana persebaran penduduk serta lahan untuk tinggal di Indonesia tidak merata. Hal ini mengakibatkan banyak masalah sosial terjadi di Indonesia salah satunya kemiskinan (Kartikasari, 2021). Kemiskinan seperti yang tertuang dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2004-2009 tidak hanya dipahami sebagai ketidakmampuan ekonomi, tetapi juga kegagalan memenuhi hak-hak dasar dan perbedaan perlakuan bagi seseorang atau sekelompok orang dalam menjalani hidupnya secara martabat. Hak-hak dasar yang diakui secara umum meliputi terpenuhinya kebutuhan pangan, kesehatan, pekerjaan, perumahan, air bersih, pertahanan, dan lingkungan hidup (Sa'dyah, 2012).

Menurut Badan Pusat Statistika (BPS), persentase penurunan kemiskinan makin berkurang dari tahun ke tahun, dimana dari tahun 1999 sampai tahun 2004 persentase kemiskinan turun sekitar 6,77% dalam 5 tahun, namun dari tahun 2014 sampai tahun 2019 penurunan kemiskinan hanya berkurang 1,74% saja yang artinya ada hambatan dari pengantas kemiskinan di Indonesia seperti pada Gambar 1. Sedangkan dari segi provinsi, menurut data BPS tahun 2013-2023, tiga provinsi yang memiliki tingkat kemiskinan tertinggi yaitu Provinsi Papua, Papua Barat, dan Nusa Tenggara Timur. Maka dari itu dibentuklah program atau kebijakan pemerintah untuk menanggulangi

kemiskinan di Indonesia salah satunya yaitu program *Sustainable Development Goals* (SDGs) dan Inpers Desa Tertinggal (IDT).



Sumber: Badan Pusat Statistika (2019)

**Gambar 1.** Perkembangan Kemiskinan di Indonesia

SDGs merupakan sebuah program pembangunan berkelanjutan dengan 17 tujuan dan 169 target pembangunan (Alisjahbana, 2018). Program ini mulai dijalankan di Indonesia sejak bulan Maret tahun 2016, dimana tujuan pertama SDGs ditargetkan untuk mengakhiri kemiskinan di mana pun (Ishartono, 2016). Selain itu, dalam Inpres Nomor 5 Tahun 1993, dijelaskan bahwa IDT adalah bagian dari gerakan nasional penanggulangan kemiskinan yang dilaksanakan secara bertahap dan berkelanjutan. Program ini memiliki beberapa tujuan yaitu pemerataan pembangunan, mengurangi kesenjangan pendapatan dan mengurangi kesenjangan laju pembangunan antar daerah, dan bantuan sosial bagi masyarakat yang paling memerlukan. Target yang ingin dicapai untuk tahun 2020-2024 adalah 6%-7%, namun tingkat kemiskinan di Indonesia pada tahun 2023 mencapai angka 9,36% yang mana masih jauh dari target yang di tetapkan.

Provinsi yang tergolong miskin adalah provinsi yang memenuhi 9 indikator rumah tangga miskin. Untuk mengetahui indiaktor rumah tangga miskin apa saja



yang ada di dalam provinsi tersebut maka dilakukan pengelompokan. Melalui pengelompokan, pemerintah dapat menyesuaikan strategi pembangunan yang sesuai dengan kebutuhan dan karakteristik masing-masing kelompok provinsi.

Indikator rumah tangga miskin di Indonesia berdasarkan laporan BPS dalam buku Penetapan Kriteria dan Variabel Pendataan Penduduk Miskin yang Komprehensif dalam Rangka Perlindungan Penduduk Miskin di Kabupaten/Kota, yaitu rumah tangga yang anggota rumah tangganya minimal lima, tingkat pendidikan kepala rumah tangga yang rendah, bekerja di sektor pertanian atau buruh dan memiliki lahan pertanian, asset rumah tangga yang kurang memadai, sumber air minum berasal dari sumur, tidak memiliki listrik, bahan bakar memasak dari kayu, mengkonsumsi daging/ayam/telur satu kali dalam seminggu, tidak ampu membayar biaya pengobatan di puskesmas, tidak memiliki modal/tabungan. Data Survei Demografi dan Kesehatan Indonesia (SDKI) merupakan merupakan survei sosial kependudukan yang dilaksanakan lima tahun sekali sejak tahun 1987, salah satu kuisisioner yang ada pada data tersebut adalah kuisisioner tentang daftar rumah tangga. Data SDKI terbaru adalah data tahun 2017, dikarenakan data survei tahun 2022 masih dalam tahap proses. Sehingga data tahun 2017 yang bisa digunakan untuk mengelompokan provinsi di Indonesia berdasarkan indikator rumah tangga miskin.

Kuisisioner rumah tangga digunakan untuk mencatat seluruh anggota rumah tangga dan keadaan rumah tangga tersebut. Isi dari kuisisioner rumah tangga meliputi tempat tinggal, jumlah anggota rumah tangga, sumber air minum, sumber air non minum, *asset* rumah tangga, jenis toilet, lokasi sumber air, jarak rumah dengan toilet, waktu pengambilan air, lokasi memasak, jenis kelamin kepala rumah tangga, bahan

lantai, bahan dinding, bahan atap, dan aset rumah tangga lainnya. Berdasarkan indikator rumah tangga miskin, variabel yang digunakan dalam penelitian ini yaitu kepemilikan mobil, sumber air minum, jenis toilet, bahan atap, bahan dinding, bahan lantai, sumber penerangan rumah tangga, bahan bakar memasak, menggunakan rekening bank, luas lantai, jumlah anggota rumah tangga, dan kepemilikan motor. Untuk mengolah data SDKI tahun 2017 menggunakan proses *data mining*.

Menurut Suntoro (2019), *data mining* adalah proses untuk mendapatkan informasi yang berguna dari data mentah yang dianalisis agar dapat membantu dalam pengambilan keputusan. Salah satu tujuan *data mining* yaitu deskriptif, digunakan untuk menemukan pola-pola yang dapat menjelaskan karakteristik dari suatu data. Teknik data mining yang termasuk dalam data mining deskriptif yaitu *clustering*. *Clustering* merupakan suatu metode pengelompokan data menjadi beberapa *cluster* di mana memiliki tingkat kemiripan yang maksimum antar data didalam satu *cluster* dan tingkat kemiripan yang minimum terhadap antar *cluster* (Nduru dkk., 2018). *Clustering* dibagi menjadi dua metode yaitu *hierarki* dan *non-hierarki*, di mana terdapat dua pendekatan dalam metode *non-hierarki* yaitu partisi dan density. Metode yang ada pada pendekatan partisi yaitu metode *K-Means*, *K-Medoids*, *K-Modes*, *Fuzzy C-Means* dan yang lainnya.

Metode *clustering* partisi yang digunakan pada penelitian ini adalah metode *K-Modes*. Metode *K-Modes* pertama kali diperkenalkan oleh Huang pada tahun 1998 yang merupakan pengembangan dari metode *K-Means* dan digunakan untuk mengelompokkan data bertipe kategorik (Huang, 2009). Terdapat beberapa

modifikasi metode *K-Modes* dari metode *K-Means* yaitu mengganti *mean* dengan modus, dan mengganti jarak *euclidean* dengan teknik *simple matching*. Yulianton (2021), menulis dalam jurnalnya beberapa kelebihan metode *K-Modes*, antara lain bisa diterapkan pada data kategorik, interpretasi yang lebih mudah, dan waktu dalam proses pembentukan *cluster* yang relatif lebih singkat, sehingga sangat cocok digunakan untuk data SDKI yang juga bertipe kategorik.

Menurut Susilowati (2020), validasi *cluster* adalah metode yang digunakan untuk melakukan evaluasi hasil dari sebuah algoritma *clustering* yang telah dilakukan dengan tujuan untuk mengetahui seberapa baik kualitas hasil *cluster*. Terdapat beberapa validasi *cluster* antara lain yaitu *Silhouette Score*, *Davies Bouildin Index* (DBI), dan *Dunn Index*. Namun validasi yang bisa digunakan untuk data kategorik adalah validasi *Davies Bouildin Index* sehingga validasi ini yang digunakan dalam penelitian. DBI diperkenalkan oleh Davied L. Davies dan Donald W. Bouildin pada tahun 1979, dimana DBI merupakan salah satu metode yang digunakan untuk mengukur validasi atau seberapa baik hasil algoritma dalam memisahkan amatan ke dalam kelompok-kelompok yang berbeda menggunakan jumlah *cluster* optimumnya (Bates dan Kalita, 2016). Metode *Davies Bouildin Index* memiliki beberapa kelebihan antara lain bisa digunakan untuk data kategorik, perhitungannya yang sederhana dan tidak memerlukan proses yang terlalu rumit, dan memperhitungkan jarak antar *centroid* sehingga lebih mudah dalam melihat kemiripan antar data.

Berdasarkan latar belakang masalah, penelitian ini membahas tentang jumlah *cluster* yang terbentuk dan karakteristik masing-masing *cluster* menggunakan algoritma *K-Modes* dengan provinsi mana saja yang perlu di utamakan dan indikator

rumah tangga miskin apa saja yang ada pada provinsi tersebut sehingga bisa digunakan untuk mengatasi masalah rumah tangga miskin di Indonesia. Penelitian ini diberi judul “Analisis *Cluster K-Modes* Dengan Validasi *Davies Bouildin Index* Dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Indikator Rumah Tangga Miskin”.

### **B. Batasan Masalah**

Berdasarkan latar belakang masalah yang telah dipaparkan, maka batasan masalah dalam penelitian ini adalah pengelompokan provinsi-provinsi di Indonesia berdasarkan indikator rumah tangga miskin menggunakan algoritma *K-Modes* dengan validasi DBI. Selain itu data yang digunakan dalam penelitian ini adalah data SDKI tahun 2017 menggunakan kuisisioner tentang rumah tangga.

### **C. Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Berapa *cluster* yang terbentuk dan Provinsi mana saja yang masuk ke dalam *cluster* tersebut menggunakan metode *K-Modes*?
2. Bagaimana karakteristik rumah tangga miskin dari *cluster* yang terbentuk pada analisis *K-Modes*?
3. Bagaimana tingkat validasi dengan jumlah *cluster* yang dibentuk pada analisis *K-Modes*?

#### **D. Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan dari penelitian ini adalah sebagai berikut.

1. Untuk mengetahui jumlah *cluster* yang terbentuk dan Provinsi yang masuk ke dalam *cluster* tersebut menggunakan metode *K-Modes*.
2. Untuk mengetahui karakteristik dari masing-masing *cluster* yang terbentuk pada analisis *K-Modes*.
3. Untuk menghitung validasi dan mengetahui seberapa baik kualitas hasil *cluster* yang digunakan dalam algoritma *K-Modes*.

#### **E. Manfaat Penelitian**

Adapun manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut.

1. Bagi peneliti, sebagai penambah ilmu dan wawasan dalam bidang clustering khususnya algoritma *K-Modes* dan validasi DBI dalam mengelompokkan rumah tangga miskin di setiap provinsi di Indonesia.
2. Bagi pembaca, sebagai bahan bacaan serta rujukan untuk melakukan penelitian berikutnya.
3. Bagi pemerintah, sebagai bahan pertimbangan atau acuan untuk membuat kebijakan dan program perencanaan pembangunan dalam menyelesaikan masalah rumah tangga miskin setiap provinsi di Indonesia.