

**PERBANDINGAN METODE NAIVE BAYES DAN  
REGRESI LOGISTIK BINER PADA KLASIFIKASI  
PENERIMA BANTUAN SOSIAL  
PROGRAM KELUARGA HARAPAN (PKH)**



**Oleh  
FANNI RAHMA SARI  
NIM. 18337043**

**DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PERNGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2023**

**PERBANDINGAN METODE NAIVE BAYES DAN  
REGRESI LOGISTIK BINER PADA KLASIFIKASI  
PENERIMA BANTUAN SOSIAL  
PROGRAM KELUARGA HARAPAN (PKH)**

**SKRIPSI**

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



**Oleh  
FANNI RAHMA SARI  
NIM. 18337043**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PERNGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2023**

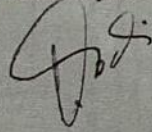
## PERSETUJUAN SKRIPSI

### PERBANDINGAN METODE NAIVE BAYES DAN REGRESI LOGISTIK BINER PADA KLASIFIKASI PENERIMA BANTUAN SOSIAL PROGRAM KELUARGA HARAPAN (PKH)

Nama : Fanni Rahma Sari  
NIM : 18337043  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

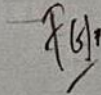
Padang, 30 Januari 2023

Mengetahui:  
Ketua Departemen Statistika



Dodi Vionanda, Ph.D  
NIP. 197806112005011002

Disetujui Oleh:  
Pembimbing



Fadhilah Fitri, S.Si., M.Stat  
NIP. 199403142022032015

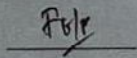
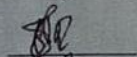
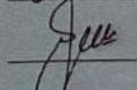
## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Fanni Rahma Sari  
NIM : 18337043  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### PERBANDINGAN METODE NAIVE BAYES DAN REGRESI LOGISTIK BINER PADA KLASIFIKASI PENERIMA BANTUAN SOSIAL PROGRAM KELUARGA HARAPAN (PKH)

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

Padang, 30 Januari 2023

Tim Penguji	Nama	Tanda Tangan
Ketua	: Fadhilah Fitri, S.Si., M.Stat	
Anggota	: Dr. Atus Amadi Putra, M.Si	
Anggota	: Dr. Dony Permana, M.Si	

## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Fanni Rahma Sari  
NIM : 18337043  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Perbandingan Metode Naive Bayes dan Regresi Logistik Biner Pada Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Ketua Departemen Statistika,



Dodi Vionanda, Ph.D  
NIP. 197806112005011002

Saya yang menyatakan,



Fanni Rahma Sari  
NIM. 18337043

**Perbandingan Metode *Naive Bayes* Dan Regresi Logistik Biner  
Pada Klasifikasi Penerima Bantuan Sosial  
Program Keluarga Harapan (PKH)**

**Fanni Rahma Sari**

**ABSTRAK**

Kepadatan penduduk menjadi salah satu penyebab dari sulitnya mencapai kesetaraan ekonomi di Indonesia. Pemerintah melakukan upaya dalam mencapai kesetaraan tersebut dengan melaksanakan berbagai program bantuan sosial. Pada tahun 2007, pemerintah membentuk suatu program bantuan sosial yang diberi nama Program Keluarga Harapan (PKH). Program ini bertujuan untuk mengatasi kemiskinan dan kelaparan melalui akses terhadap pendidikan dan kesehatan. Permasalahan yang sering terjadi dalam penyaluran bantuan sosial adalah penerimaan yang tidak tepat sasaran. Hal tersebut dibuktikan oleh Badan Pemeriksa Keuangan (BPK) yang menemukan kesalahan penyaluran bansos dari pemerintah sehingga mengakibatkan kerugian negara mencapai Rp 6,9 triliun. Oleh karena itu, diperlukan metode yang dapat mengukur tingkat klasifikasi data penerima bantuan sosial dengan tepat dan akurat yaitu dengan menggunakan metode *Naive Bayes* dan Regresi Logistik. Kedua metode sangat tepat digunakan karena keduanya sama-sama baik dalam pengklasifikasian data yang berjumlah besar.

Penelitian ini menggunakan data sekunder yang diperoleh langsung dari Kantor Lurah Anduring. Data yang digunakan adalah data DTKS bantuan sosial PKH tahun 2020. DTKS merupakan Data Terpadu Kesejahteraan Sosial yang digunakan pemerintah untuk menyalurkan bantuan sosial. Data dibagi menjadi data *training* dan data *testing* dengan masing-masing 80% dan 20%.

Hasil yang diperoleh dengan menggunakan metode *Naive Bayes* dan Regresi Logistik Biner yaitu bahwa dengan menggunakan data *testing* sebesar 20% diperoleh hasil akurasi dengan metode Regresi Logistik Biner sebesar 75% dan eror 25%. Sedangkan metode *Naive Bayes* memiliki hasil akurasi sebesar 70% dan eror 30%. Dari hasil tersebut disimpulkan bahwa metode Regresi Logistik Biner adalah metode terbaik dalam mengukur ketepatan klasifikasi penerima bantuan sosial PKH di Kelurahan Anduring tahun 2020.

Kata Kunci: Bantuan Sosial, PKH, Klasifikasi, Regresi Logistik Biner, *Naive Bayes*

# **Comparison of Naive Bayes Method and Binary Logistics Regression on the Classification of Recipients of Social Assistance for the Program Keluarga Harapan (PKH)**

**Fanni Rahma Sari**

## **ABSTRACT**

Population density is one of the causes of the difficulty of achieving economic equality in Indonesia. The government is making efforts to achieve this equality by implementing various social assistance programs. In 2007, the government established a social assistance program called the Program Keluarga Harapan (PKH). This program aims to overcome poverty and hunger through access to education and health. The problem that often occurs in the distribution of social assistance is receipt that is not right on target. This was proven by the Badan Pemeriksa Keuangan (BPK) which found an error in distributing social assistance from the government, resulting in state losses reaching 6.9 trillion rupiah. Therefore, a method is needed that can measure the level of classification of social assistance recipient data accurately and precisely, namely by using the Naive Bayes and Logistics Regression. Both methods are very appropriate to use because both are equally good at classifying large amounts of data.

This study uses secondary data obtained directly from the Anduring Lurah Office. The data used is the 2020 PKH social assistance DTKS data. DTKS is the Data Terpadu Kesejahteraan Sosial used by the government to distribute social assistance. data is divided into training data and testing with 80% and 20% respectively.

The results obtained using the Naive Bayes and Binary Logistics Regression are that by using testing of 20%, the accuracy results with the Binary Logistics Regression method are 75% and the error is 25%. While the Naive Bayes has an accuracy of 70% and an error of 30%. From these results, it is concluded that the Binary Logistics Regression method is the best method in measuring the accuracy of the classification of PKH social assistance recipients in Anduring Village in 2020.

**Keywords:** Social Assistance, PKH, Classification, Binary Logistics Regression, Naive Bayes

## KATA PENGANTAR

Alhamdulillah, puji dan syukur tiada hentinya penulis ucapkan kepada Allah SWT atas segala pertolongan, rahmat, nikmat dan kasih sayang-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Perbandingan Metode *Naive Bayes* Dan Regresi Logistik Biner Pada Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan (PKH)”**.

Penyusunan skripsi ini untuk memenuhi persyaratan memperoleh gelar Sarjana pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universtas Negeri Padang. Penyusunan skripsi ini tidak terlepas dari bantuan dan kerjasama berbagai pihak. Termakasih penulis sampaikan kepada:

1. Ibu Dina Fitria, S.Pd., M.Si sebagai Dosen Pembimbing Akademik.
2. Ibu Fadhilah Fitri, S.Si., M.Stat sebagai Dosen Pembimbing Skripsi.
3. Bapak Dr. Dony Permana, M.Si sebagai Dosen Penguji dalam seminar proposal penelitian dan sidang skripsi penulis.
4. Bapak Drs. Atus Amadi Putra, M.Si sebagai Dosen Penguji dalam seminar proposal penelitian dan sidang skripsi penulis.
5. Bapak Dodi Vionanda, Ph.D. sebagai Kepala Departemen Statistika, serta Kepala Prodi Sarjana Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.
6. Bapak dan Ibu Dosen serta Staf Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah membantu penulis selama menimba ilmu di Program Studi Statistika.



7. Teristimewa kepada kedua orang tua, abang, dan kakak yang telah memberikan semangat, doa, dan dukungan selama proses penyelesaian skripsi hingga saat ini.
8. Teman-teman statistika angkatan 2018 yang telah memberikan semangat dan mendukung penulis dalam pembuatan skripsi ini.
9. Semua pihak terkait yang telah ikut berkontribusi dalam penulisan skripsi ini.

Semoga skripsi ini kedepannya dapat bermanfaat bagi penulis dan semua pihak. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan masukan dan saran dari pembaca. Akhir kata penulis ucapkan terima kasih.

Padang, Januari 2023

Penulis

## DAFTAR ISI

ABSTRAK .....	iv
KATA PENGANTAR.....	vi
DAFTAR ISI .....	viii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN .....	xi
BAB I PENDAHULUAN .....	1
A. Latar Belakang.....	1
B. Batasan Masalah .....	8
C. Rumusan Masalah.....	8
D. Tujuan Penelitian .....	8
E. Manfaat Penelitian .....	9
BAB II KERANGKA TEORITIS .....	10
A. Bantuan Sosial Program Keluarga Harapan (PKH).....	10
B. Kelurahan.....	14
C. Klasifikasi .....	16
D. Regresi Logistik Biner .....	17
E. Naive Bayes .....	26
F. Variabel Dummy .....	28
G. Data Mining .....	28
H. Confussion Matrix .....	29
BAB III METODOLOGI PENELITIAN .....	31
A. Sumber Data .....	31
B. Variabel Penelitian .....	31
C. Langkah-Langkah Analisis Data .....	32
BAB IV HASIL DAN PEMBAHASAN.....	35
A. Statistika Deskriptif .....	35
B. Analisis Data.....	51
BAB V PENUTUP .....	64
A. Kesimpulan .....	64
B. Saran .....	64
DAFTAR PUSTAKA.....	66
LAMPIRAN .....	69

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. <i>Confussion Matrix</i> .....	29
2. Variabel Dependen .....	31
3. Variabel Independen.....	31
4. Pembagian Data <i>Training</i> dan Data <i>Testing</i> .....	51
5. Uji Wald .....	52
6. Uji Kecocokan Model .....	53
7. Pemilihan Model Terbaik .....	53
8. <i>Odds Ratio</i> .....	54
9. <i>Confussion Matrix</i> Regresi Logistik Biner.....	55
10. Probabilitas Penerima Bantuan PKH .....	56
11. Penerima Bantuan dengan Status Bangunan .....	57
12. Probabilitas Penerima Bantuan dengan Status Bangunan .....	57
13. Probabilitas Penerima Bantuan dengan Kondisi Dinding .....	57
14. Probabilitas Penerima Bantuan dengan Kondisi Atap .....	57
15. Probabilitas Penerima Bantuan dengan Sumber Air Minum .....	58
16. Probabilitas Penerima Bantuan dengan Sumber Penerangan.....	58
17. Probabilitas Penerima Bantuan dengan Bahan Bakar Masak .....	58
18. Probabilitas Penerima Bantuan dengan Ketersediaan Tabung Gas.....	58
19. Probabilitas Penerima Bantuan dengan Ketersediaan Lemari Es .....	59
20. Probabilitas Penerima Bantuan dengan Ketersediaan AC .....	59
21. Probabilitas Penerima Bantuan dengan Ketersediaan Telepon.....	59
22. Probabilitas Penerima Bantuan dengan Ketersediaan Televisi .....	59
23. Probabilitas Penerima Bantuan dengan Ketersediaan Emas .....	59
24. Probabilitas Penerima Bantuan dengan Ketersediaan Motor .....	60
25. Probabilitas Penerima Bantuan dengan Ketersediaan Mobil .....	60
26. Probabilitas Penerima Bantuan dengan Ketersediaan Aset Tak Bergerak.....	60
27. <i>Confussion Matrix Naive Bayes</i> .....	62
28. Perbandingan Klasifikasi Regresi Logistik Biner dan <i>Naive Bayes</i> .....	63

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Jumlah Penerima PKH Tahun 2007 s.d. 2020.....	11
2. Total Alokasi Anggaran PKH Tahun 2007 s.d. 2020.....	12
3. Rata-Rata Besarnya Bantuan yang diterima per orang Tahun 2007 s.d. 2020.....	12
4. Letak Kelurahan Anduring di Kecamatan Kuranji.....	16
5. Kurva Fungsi Logistik.....	20
6. <i>Flowchart</i> Tahapan Analisis Data.....	34
7. Perbandingan Penerima Bantuan PKH.....	35
8. Penerima Bantuan Sosial PKH berdasarkan Status Bangunan .....	36
9. Penerima Bantuan Sosial PKH berdasarkan Kondisi Dinding.....	37
10. Penerima Bantuan Sosial PKH berdasarkan Kondisi Atap .....	38
11. Penerima Bantuan Sosial PKH berdasarkan Sumber Air Minum .....	39
12. Penerima Bantuan Sosial PKH berdasarkan Sumber Penerangan .....	40
13. Penerima Bantuan Sosial PKH berdasarkan Bahan Bakar Masak .....	41
14. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Tabung Gas .....	42
15. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Lemari Es .....	43
16. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan AC .....	44
17. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Telepon.....	45
18. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Televisi.....	46
19. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Emas.....	47
20. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Motor.....	48
21. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Mobil.....	49
22. Penerima Bantuan Sosial PKH berdasarkan Ketersediaan Aset Tak Bergerak .....	50

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Output R Studio metode Regresi Logistik Biner.....	69
2. Output R Studio metode Naive Bayes.....	71

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang**

Peningkatan perlindungan dan kesejahteraan sosial merupakan salah satu prioritas pemerintah terhadap mereka yang termasuk ke dalam kelompok masyarakat kurang mampu. Perlindungan dan kesejahteraan sosial di Indonesia ini diwujudkan dalam bentuk bantuan sosial dan jaminan sosial (Avriliyanti, 2015). Pemerintah berharap kesejahteraan masyarakat di Indonesia bisa mendapatkan keadilan yang sama rata dalam kehidupan berbangsa dan bernegara untuk menciptakan masyarakat yg makmur, adil dan beradab sesuai dengan yang tertulis dalam alinea keempat pembukaan Undang-Undang Dasar 1945.

Berdasarkan situs *Worldometer* (2020), Indonesia menempati peringkat ke-4 dengan jumlah penduduk terbesar di dunia yaitu sebesar 273.523.615. Hal ini terjadi karena pertumbuhan penduduk yang sangat cepat mencapai angka 1,25% ditahun tersebut. Melihat hal tersebut, akan sulit untuk mendapatkan kesetaraan kehidupan masyarakat di Indonesia nantinya. Menurut Undang-Undang Nomor 11 Tahun 2009, negara bertanggung jawab atas kesejahteraan rakyatnya. Oleh karena itu, pemerintah terus melakukan upaya dalam mencapai kesetaraan kehidupan baik di daerah perkotaan maupun pedesaan dengan melaksanakan berbagai program bantuan sosial.

Bantuan sosial adalah pemberian bantuan secara selektif dalam bentuk uang atau barang kepada masyarakat yang bertujuan untuk meningkatkan kesejahteraan masyarakat (Peraturan Menteri Sosial No 1 Tahun 2019). Penyediaan anggaran

dana bantuan sosial harus dijabarkan secara rinci agar penerimaan dan sasaran penggunaannya jelas serta tepat sasaran. Pertanggungjawaban yang baik juga sangat diperlukan dalam pengelolaannya karena berhubungan dengan pendapatan, pengeluaran, dan pemberian dana sesuai dengan prosedur dalam Undang-Undang Dasar 1945.

Pada tahun 2007 pemerintah membuat suatu program bantuan sosial yang diberi nama Program Keluarga Harapan (PKH) dengan tujuan memutus rantai kemiskinan dan meningkatkan kesejahteraan serta kualitas sumber daya manusia (Nataya & Supriyadi, 2017). Menurut Kementerian Sosial (Kemensos), PKH adalah salah satu program pemberian bantuan sosial bersyarat kepada Keluarga Miskin (KM) yang telah ditetapkan sebagai keluarga penerima. Tujuan dari PKH yaitu mengatasi penduduk miskin dan kelaparan melalui akses terhadap pendidikan dan kesehatan. Besar bantuan yang didapatkan bisa berubah sesuai dengan kondisi keluarga saat itu atau apabila penerima tidak dapat memenuhi syarat yang ditentukan lagi (Rizki, 2019). Menurut Kemensos, bantuan PKH dibagikan berdasarkan Data Terpadu Kesejahteraan Sosial (DTKS) yaitu basis data yang digunakan pemerintah untuk menyalurkan bantuan sosial tersebut. Pendataan dilakukan secara berkala minimal satu tahun sekali oleh pemerintah daerah Kabupaten atau Kota. Selanjutnya hasil pendataan disampaikan langsung kepada pemerintah daerah provinsi untuk diteruskan kepada Kementerian Sosial.

Kota Padang menjadi salah satu daerah yang sudah menyalurkan bantuan PKH di Sumatera Barat. Penerima bantuan ini akan di data terlebih dahulu oleh pihak kecamatan melalui masing-masing kelurahan. Kecamatan Kuranji sebagai

salah satu kecamatan di Kota Padang yang memiliki 9 kelurahan, salah satunya Kelurahan Anduring. Pada akhir tahun 2021, Kelurahan Anduring ditunjuk sebagai “Kelurahan Statistik” yang ada di Kecamatan Kuranji. Hal ini dilihat dari ketersediaan data yang lengkap serta keakuratan data yang dimiliki terbilang cukup tinggi. Berdasarkan hasil observasi yang telah dilakukan di Kelurahan Anduring, daftar nama penerima bantuan sosial PKH sudah dikeluarkan melalui data DTKS. Daftar nama penerima ini dikeluarkan melalui tahapan pendataan, verifikasi, dan validasi data secara mandiri oleh Kelurahan Anduring.

Pada tahun 2022 Badan Pemeriksa Keuangan (BPK) menemukan kesalahan penyaluran bansos dari pemerintah yang mengakibatkan kerugian negara mencapai Rp 6,9 triliun. Laporan Hasil Pemeriksaan (LHP) Semester II Tahun 2021 menyebut kesalahan penyaluran bansos terjadi pada Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), Bantuan Sosial tunai (BST). Dalam laporannya, BPK menyebutkan ada enam kesalahan penyaluran bansos yang tidak sesuai dengan ketentuan sehingga penerima manfaat tidak tepat sasaran. Hal ini juga dibuktikan melalui penelitian yang telah dilakukan oleh Saiful Mujani (2020) yaitu sebanyak 54% responden menyatakan program bantuan sosial kurang/tidak tepat sasaran, 37% responden menyatakan program bantuan sudah tepat sasaran dan 8% responden menjawab tidak tahu. Oleh karena itu, diperlukan metode yang dapat mengukur tingkat klasifikasi data penerima bantuan sosial dengan tepat dan akurat. Tujuannya yaitu untuk mempermudah pihak yang berwenang dalam membagikan bantuan agar tepat sasaran dan sampai kepada masyarakat sesuai yang diharapkan.



Proposal ini akan membahas mengenai proses penentuan kelayakan penerima dengan mengetahui tingkat ketepatan klasifikasi penerima bantuan sosial PKH di Kelurahan Anduring. Berdasarkan penelitian terdahulu yang dilakukan oleh Rahmansyah, dkk (2020), menyimpulkan bahwa data yang digunakan seringkali tidak akurat sehingga menimbulkan banyak permasalahan yang sering terjadi dalam penyaluran bantuan sosial. Beberapa masyarakat yang berhak menerima bisa jadi tidak menerima dan masyarakat yang bukan dikategorikan berhak justru menerima bantuan tersebut. Hal tersebut diperkuat dengan keterangan yang diperoleh dari pihak kelurahan bahwa ada beberapa warga yang datang ke kantor lurah untuk menanyakan terkait bantuan yang belum juga mereka dapatkan, pihak kelurahan menjelaskan bahwa data penerima bantuan PKH tersebut dikeluarkan oleh DTKS dan tidak dari kelurahan. Melihat hal tersebut diperlukan suatu sistem yang dapat membantu pemerintah dalam melakukan pembaruan data dengan harapan agar kebijakan yang dikeluarkan dapat berjalan dengan efektif. Jika penerima bantuan dan kelayakannya terdata dengan baik, maka akan mengurangi ketidaksinkronan penerima bantuan dan akan lebih memudahkan pihak berwenang dalam pembagiannya. Oleh karena itu, diperlukan metode yang tepat untuk menangani hal tersebut.

Dalam analisis statistika ada beberapa metode yang digunakan dalam klasifikasi data diantaranya yaitu Analisis Diskriminan, Regresi Logistik, *Naive Bayes*, dan lain-lain. Klasifikasi adalah metode untuk mengelompokkan suatu data yang disusun secara sistematis. *Naive Bayes* merupakan algoritma klasifikasi sederhana dan mudah diimplementasikan sehingga algoritmanya sangat efektif

apabila diuji dengan dataset yang tepat terutama dengan seleksi fitur, *Naive Bayes* akan dapat mengurangi terjadinya data berganda dalam satu tempat yang disebut *redundant* pada data (Witten, dkk, 2011). Klasifikasi *Naive Bayes* adalah metode klasifikasi yang berdasarkan pada teori probabilitas dan teorema bayes (Norfriansyah, 2016). Menurut Han and Kamber (2006) klasifikasi *Naive Bayes* terbukti memiliki akurasi dan ketepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar. Keuntungan menggunakan *Naive Bayes* adalah metode ini hanya membutuhkan jumlah *training data* yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Patterkari dan Parveen, 2012).

Selanjutnya, juga dapat dilakukan dengan metode Regresi Logistik Biner. Klasifikasi dengan menggunakan Regresi Logistik adalah suatu metode pengklasifikasian dengan setiap pengamatan memiliki variabel prediktor yang berkaitan dengan variabel respon. Metode Regresi Logistik Biner adalah salah satu metode yang dapat digunakan untuk menyelesaikan berbagai jenis permasalahan. Metode ini mampu menjelaskan faktor-faktor yang mempengaruhi variabel dependen dan pengklasifikasian dari suatu data (Samosir, 2015). Menurut Kuncoro (2001) kelebihan lainnya dari regresi logistik adalah tidak memiliki asumsi normalitas dan heteroskedastisitas atas variabel bebasnya sehingga tidak diperlukan uji asumsi klasik walaupun menggunakan variabel independen yang berjumlah lebih dari satu. Variabel independen dalam regresi logistik bisa berupa variabel kontinu, distrik, dan dikotomi. Regresi logistik tidak membutuhkan keterbatasan variabel independen dan tidak mengharuskan variabel bebasnya dalam bentuk

interval. Oleh karena itu, penelitian ini akan menggunakan dua metode analisis tersebut untuk membandingkan metode mana yang lebih baik pada data yang digunakan.

Klasifikasi pada data dengan jumlah banyak dan atribut yang beragam sering menghasilkan tingkat akurasi menjadi rendah. Untuk itu diperlukan metode yang dapat menangani hal tersebut yaitu metode *Naive Bayes* dan Regresi Logistik (Salim, 2017). Persamaan dari metode *Naive Bayes* dan Regresi Logistik yaitu keduanya sama-sama memiliki ketepatan klasifikasi yang baik dibandingkan dengan metode lainnya dan keduanya juga tepat digunakan untuk menangani klasifikasi data dalam jumlah besar yang sering menghasilkan tingkat akurasi rendah. Metode *Naive Bayes* juga mudah dan cepat diaplikasikan, bisa menyatu lebih cepat dengan Regresi Logistik, algoritma klasifikasinya juga dapat digunakan untuk biner maupun multinomial. *Naive Bayes* memiliki kelebihan yaitu tidak memerlukan uji statistik seperti uji signifikansi (Putri dkk, 2014). Hal tersebut menyebabkan tidak dapat melihat apakah persamaan yang didapat valid untuk memprediksi variabel dependen atau tidak. Sedangkan metode Regresi Logistik Biner memiliki beberapa uji prasyarat atau kualitas data yang dilakukan contohnya yaitu uji kelayakan model. Pada *Naive Bayes* jika variabel independen terlalu banyak maka variabel dependen akan sulit untuk diprediksi sehingga tingkat akurasi berkurang atau menjadi kecil tetapi pada Regresi Logistik tidak memiliki keterbatasan data pada variabel independen dalam memprediksi tingkat akurasi.

Penelitian sebelumnya telah dilakukan oleh Parthasarathy dan Madhu (2020) yang membandingkan antara metode *Naive Bayes* dan Regresi Logistik. Penelitian ini dilakukan dengan menggunakan data medis melalui pendekatan *Machine Learning*. Model Regresi Logistik dan *Naive Bayes* menunjukkan nilai akurasi pada *testing set* 25% masing-masing yaitu 92,3 dan 86,5. Demikian pula dengan meningkatkan *testing set* menjadi 40% menunjukkan hasil yaitu 92,6 dan 81,7. Jika dibandingkan dengan model *Naive Bayes*, model Regresi Logistik memberikan akurasi terbaik dan ketika ukuran *testing set* ditingkatkan, model Regresi Logistik masih berkerja lebih baik daripada model *Naive Bayes*. Hal berbeda ditunjukkan dari penelitian Fajrila (2018) dengan membandingkan metode Regresi Logistik Biner dan *Naive Bayes Classifier* pada klasifikasi ketepatan waktu kelulusan mahasiswa. Hasil yang diperlihatkan adalah algoritma *Naive Bayes Classifier* lebih baik dalam mengklasifikasi lama studi mahasiswa karena akurasi data uji sebesar 89,10% dibandingkan dengan akurasi data uji Regresi Logistik Biner sebesar 82,29%.

Penelitian ini cocok diolah dengan menggunakan kedua metode karena memiliki variabel dependen yang bersifat dikotomi (dua variabel) serta variabel independent bersifat kategorik dan memiliki sampel dalam jumlah besar sehingga tepat digunakan dengan metode Regresi Logistik Biner. Pada metode *Naive Bayes*, data ini cocok digunakan karena antar satu kelas dengan kelas lainnya tidak saling bergantung (independen).

Berdasarkan pemaparan yang sudah dijelaskan sebelumnya, disimpulkan bahwa diperlukan analisis untuk mengetahui tingkat akurasi dari klasifikasi

penerima bantuan sosial PKH dengan menggunakan metode *Naive Bayes* dan Regresi Logistik Biner sehingga dapat meminimalisir terjadinya kesalahan dalam pembagiannya. Maka dari itu, dilakukan penelitian yang berjudul “**Perbandingan Metode *Naive Bayes* dan Regresi Logistik Biner Pada Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan (PKH)**”.

#### **B. Batasan Masalah**

Batasan-batasan masalah bertujuan agar pembahasan dari penelitian ini tidak melebar sehingga keluar dari pokok-pokok permasalahan yang sudah dirumuskan sebelumnya, maka dari itu batasan masalah dari penelitian ini sebagai berikut :

1. Data yang digunakan adalah Data Terpadu Kesejahteraan Sosial (DTKS) di Kelurahan Anduring tahun 2020
2. Analisis yang digunakan adalah *Naive Bayes* dan Regresi Logistik Biner

#### **C. Rumusan Masalah**

Berdasarkan latar belakang tersebut, rumusan masalah dalam proposal penelitian ini sebagai berikut:

1. Bagaimana ketepatan klasifikasi penerima bantuan sosial PKH dengan menggunakan metode Regresi Logistik Biner?
2. Bagaimana ketepatan klasifikasi penerima bantuan sosial PKH dengan menggunakan metode *Naive Bayes*?
3. Bagaimana perbandingan tingkat akurasi dari ketepatan klasifikasi dengan metode Regresi Logistik Biner dan *Naive Bayes*?

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

Berdasarkan rumusan masalah penelitian, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui klasifikasi penerima bantuan sosial PKH dengan metode Regresi Logistik Biner
2. Untuk mengetahui klasifikasi penerima bantuan sosial PKH dengan metode *Naive Bayes*
3. Untuk mengetahui perbandingan tingkat akurasi dari klasifikasi Regresi Logistik Biner dan *Naive Bayes*

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

Manfaat dari penelitian adalah sebagai berikut:

1. Bagi peneliti, penelitian ini bermanfaat untuk menambah wawasan dan pengalaman dalam analisis data dengan menerapkan teori-teori yang sudah dipelajari
2. Bagi masyarakat, penelitian ini diharapkan bisa memberikan informasi pengetahuan agar kedepannya bisa lebih baik
3. Bagi pembaca, penelitian ini bermanfaat untuk menambah pengetahuan mengenai kelayakan penerima bantuan PKH dengan mengetahui tingkat ketepatan klasifikasi bagi penerima bantuan serta dapat dijadikan pedoman untuk penelitian-penelitian selanjutnya