

**PENERAPAN *INFORMATION GAIN* UNTUK  
SELEKSI ATRIBUT PADA KLASIFIKASI  
PENERIMA BANTUAN SOSIAL  
MENGUNAKAN *NAÏVE BAYES***

**(Studi Kasus : Desa Muara Bone)**

**Oleh**

**FAHRUN LAKADJO**

**T3120017**

**SKRIPSI**

**Untuk memenuhi salah satu syarat ujian  
guna memperoleh gelar Sarjana**



**PROGRAM SARJANA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS ICHSAN GORONTALO  
GORONTALO  
2024**

## **PERSETUJUAN SKRIPSI**

# **PENERAPAN *INFORMATION GAIN* UNTUK SELEKSI ATRIBUT PADA KLASIFIKASI PENERIMA BANTUAN SOSIAL MENGUNAKAN *NAÏVE BAYES***

**(Studi Kasus : Desa Muara Bone)**

Oleh

**FAHRUN LAKADJO**

**T3120017**

## **SKRIPSI**

Untuk memnuhi salah satu syarat ujian  
guna memperoleh gelar Sarjana  
Program Studi Teknik Informatika,  
Ini telah disetujui oleh Tim Pembimbing  
Gorontalo, 04 Juni 2024

**Pembimbing Utama**



**Amiruddin, M.Kom**  
**NIDN. 0910097601**

**Pembimbing Pendamping**



**Warid Yunus, M.Kom**  
**NIDN. 0914059001**



## PENGESAHAN SKRIPSI

# PENERAPAN *INFORMATION GAIN* UNTUK SELEKSI ATRIBUT PADA KLASIFIKASI PENERIMA BANTUAN SOSIAL MENGUNAKAN *NAÏVE BAYES*

(Studi Kasus : Desa Muara Bone)

Oleh

FAHRUN LAKADJO

T3120017

Diperiksa oleh Panitia Ujian Strata Satu (S1)  
Universitas Ichsan Gorontalo

1. Ketua Penguji

Haditsah Annur, M.Kom

2. Anggota

Husdi, M.Kom

3. Anggota

Misrawaty A. Puspa, M.Kom

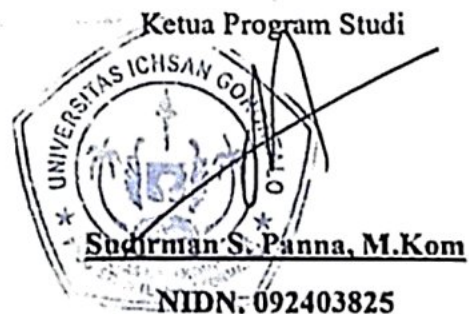
4. Anggota

Amiruddin, M.Kom

5. Anggota

Warid Yunus, M.Kom

Mengetahui



## PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, 4 Juni 2024

Yang Membuat Pernyataan,

A handwritten signature in black ink is placed over a purple revenue stamp. The stamp features the Garuda Pancasila emblem and the text '10000', 'METERAI TEMPEL', and a unique alphanumeric code 'CD3ALX214486528'.

**Fahrur Lakadjo**

**NIM: T3120017**

## ABSTRACT

**FAHRUN LAKADJO. T3120017. THE APPLICATION OF INFORMATION GAIN FOR ATTRIBUTE SELECTION IN THE CLASSIFICATION OF SOCIALASSISTANCE RECIPIENTS USING NAÏVE BAYES**

*This research discusses the attribute selection significance in the social assistance recipients classification, using the Naïve Bayes method by applying Information Gain as an attribute selection technique. Naïve Bayes, although popular due to its simplicity and efficiency, has a significant drawback: the assumption of independence between features. This assumption is often unrealistic in many practical cases, which can reduce prediction accuracy. To cope with the weakness, attribute selection is used to identify and use only the most relevant features, thereby improving model performance. In this study, the data employed cover eleven attributes, namely Household Occupation, Wife's Occupation, Household Income, Wife's Income, Number of Dependents, House Ownership, Electricity Source, Land Area, Building Area, Water Source, and Toilets. Through Information Gain analysis, it explains that seven (7) attributes (Household Income, Wife's Income, Household Occupation, Land Area, Building Area, Toilets, and Wife's Occupation) significantly affect classification. The experimental results show that the accuracy of the Naïve Bayes model before attribute selection is 85%, while after attribute selection using Information Gain, the accuracy increases to 96%. The 11% increase in accuracy shows that attribute selection with Information Gain can improve classification performance, making the model more accurate and reliable in determining social assistance recipients. This finding confirms the importance of attribute selection in the improvement of the classification performance models using Naïve Bayes.*

*Keywords: information gain, attribute selection, classification, Naive Bayes, socialassista*



## ABSTRAK

### **FAHRUN LAKADJO. T3120017. PENERAPAN *INFORMATION GAIN* UNTUK SELEKSI ATRIBUT PADA KLASIFIKASI PENERIMA BANTUANSOSIAL MENGGUNAKAN *NAÏVE BAYES***

Penelitian ini membahas pentingnya seleksi atribut dalam klasifikasi penerima bantuan sosial menggunakan metode *Naïve Bayes*, dengan menerapkan *Information Gain* sebagai teknik seleksi atribut. *Naïve Bayes*, meskipun populer karena kesederhanaan dan efisiensinya, memiliki kelemahan signifikan: asumsi independensi antar fitur. Asumsi ini seringkali tidak realistis dalam banyak kasus praktis, sehingga dapat mengurangi akurasi prediksi. Untuk mengatasi kelemahan ini, seleksi atribut digunakan untuk mengidentifikasi dan menggunakan hanya fitur-fitur yang paling relevan, sehingga dapat meningkatkan kinerja model. Dalam penelitian ini, data yang digunakan mencakup 11 atribut: Pekerjaan KK, Pekerjaan Istri, Penghasilan KK, Penghasilan Istri, Jumlah Tanggungan, Kepemilikan Rumah, Sumber Listrik, Luas Tanah, Luas Bangunan, Sumber Air, dan MCK. Melalui analisis *Information Gain*, ditemukan bahwa 7 atribut (Penghasilan KK, Penghasilan Istri, Pekerjaan KK, Luas Tanah, Luas Bangunan, MCK, dan Pekerjaan Istri) memiliki pengaruh signifikan terhadap klasifikasi. Hasil eksperimen menunjukkan bahwa akurasi model *Naïve Bayes* sebelum seleksi atribut adalah 85%, sedangkan setelah seleksi atribut menggunakan *Information Gain*, akurasi meningkat menjadi 96%. Peningkatan akurasi sebesar 11% ini menunjukkan bahwa seleksi atribut dengan *Information Gain* mampu meningkatkan kinerja klasifikasi, menjadikan model lebih akurat dan andal dalam menentukan penerima bantuan sosial. Temuan ini menegaskan pentingnya seleksi atribut dalam memperbaiki performa model klasifikasi yang menggunakan *Naïve Bayes*.

Kata kunci: *Information Gain*, seleksi atribut, klasifikasi, *Naive Bayes*, bantuansosial.

## KATA PENGANTAR

Alhamdulillah, Penulis dapat menyelesaikan skripsi ini dengan judul : **PENERAPAN *INFORMATION GAIN* UNTUK SELEKSI ATRIBUT PADA KLASIFIKASI PENERIMA BANTUAN SOSIAL MENGGUNAKAN *NAÏVE BAYES***, sebagai salah satu syarat Ujian Akhir guna memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Ibu Dr. Hj Juriko Abdusamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abd Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan A Salihi, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Bapak Amiruddin, M.Kom, selaku Pembimbing I yang telah banyak membantu peneliti dalam menyelesaikan skripsi ini.
8. Bapak Warid Yunus, M.Kom, selaku Pembimbing II yang telah banyak membantu peneliti dalam menyelesaikan skripsi ini.

9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan membimbing serta mengajarkan berbagai disiplin ilmu kepada penulis;
10. Ucapan terima kasih kepada Kedua Orang Tua Saya yang tercinta, atas segala kasih sayang, Jerih payah, Motivasi dan Doa restunya kepada penulis.
11. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
12. Kepada semua pihak yang ikut membantu dalam penyelesaian usulan penelitian ini yang tak sempat penulis sebutkan satu-persatu;

Dengan demikian, penulis menyadari bahwa masih banyak terdapat kekurangan dalam penyusunan skripsi. Maka dari itu, diharapkan saran dan kritik yang konstruktif. Harapannya semoga hasil yang sudah dicapai bisa bermanfaat untuk kita semua, Aamiin.

Gorontalo, 4 Juni 2024

**Penulis**

## DAFTAR ISI

<b>HALAMAN SAMPUL .....</b>	<b>i</b>
<b>PERSETUJUAN SKRIPSI.....</b>	<b>ii</b>
<b>PENGESAHAN SKRIPSI.....</b>	<b>iii</b>
<b>PERNYATAAN SKRIPSI.....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>ix</b>
<b>DAFTAR GAMBAR .....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>DAFTAR PSEUDOCODE.....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	4
1.3 Rumusan Masalah .....	4
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian.....	5
<b>BAB II LANDASAN TEORI.....</b>	<b>6</b>
2.1 Tinjauan Studi Kasus.....	6
2.2 Tinjauan Pustaka.....	8
2.2.1 Bantuan Sosisal .....	8
2.2.2 Data Mining .....	9
2.2.3 <i>Naïve Bayes</i> .....	12
2.2.4 Confusion Matrix .....	17
2.3 Kerangka Pikir.....	20
<b>BAB III METODE PENELITIAN .....</b>	<b>21</b>

3.1	Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian .....	21
3.2	Pengumpulan Data .....	21
3.2.1	Observasi.....	23
3.2.2	Wawancara .....	23
3.2.3	Dokumentasi .....	23
3.3	Pengembangan Model .....	23
3.4	Evaluasi Model.....	24
3.5	Analisis Deskriptif.....	24
<b>BAB IV HASIL PENELITIAN.....</b>		<b>25</b>
4.1	Hasil Pengumpulan Data .....	25
4.2	Hasil Pemodelan .....	26
4.2.1	Data Preprocessing .....	27
4.2.2	Pemodelan dengan Metode <i>Information Gain</i> .....	32
4.2.3	Pemodelan dengan <i>Naive Bayes</i> .....	35
4.2.4	Pengujian Model Dan Evaluasi Kinerja Model .....	39
4.2.5	Implementasi Hasil Pemodelan.....	40
4.2.6	Implementasi Data Processing .....	40
<b>BAB V PEMBAHASAN PENELITIAN.....</b>		<b>51</b>
5.1	Pembahasan Kinerja Model.....	51
5.2	Pembahasan Hasil Pemodelan .....	52
5.2.1	Hasil Pemodelan <i>Naïve Bayes</i> Sebelum Seleksi Atribut.....	52
5.2.2	Hasil Seleksi Atribut Penting dengan <i>Information Gain</i> .....	55
5.2.3	Hasil Pemodelan <i>Naive Bayes</i> Setelah Seleksi Atribut .....	56
<b>BAB VI KESIMPULAN DAN SARAN .....</b>		<b>60</b>
6.1	Kesimpulan.....	60
6.2	Saran .....	60



<b>DAFTAR PUSTAKA.....</b>	<b>61</b>
----------------------------	-----------

## DAFTAR GAMBAR

Gambar 2. 1 Proses KDD.....	10
Gambar 2. 2 Alur Metode <i>Naive Bayes</i> .....	15
Gambar 2. 3 Kerangka Pikir.....	20
Gambar 5. 1 Hasil Perhitungan <i>Confusion Matrix</i> .....	53
Gambar 5. 2 Hasil Skor Kepentingan .....	55
Gambar 5. 3 Hasil Perhitungan <i>Confusion Matrix</i> .....	57

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait.....	6
Tabel 2. 2 <i>Confusion Matrix</i> .....	17
Tabel 3. 1 Atribut Data Penerima Bantuan Langsung Tunai .....	21
Tabel 4. 1 Hasil Pengumpulan Data Set.....	25
Tabel 4. 2 Konversi Atribut Pekerjaan KK .....	27
Tabel 4. 3 Konversi Atribut Pekerjaan Istri.....	27
Tabel 4. 4 Konversi Atribut Penghasilan KK.....	28
Tabel 4. 5 Konversi Atribut Penghasilan Istri .....	28
Tabel 4. 6 Konversi Atribut Jumlah Tanggungan.....	28
Tabel 4. 7 Konversi Atribut Luas Tanah .....	29
Tabel 4. 8 Konversi Atribut Luas Bangunan .....	29
Tabel 4. 9 Konversi Atribut Kepemilikan Rumah.....	29
Tabel 4. 10 Konversi Atribut Sumber Listrik.....	30
Tabel 4. 11 Konversi Atribut Sumber Air .....	30
Tabel 4. 12 Konversi Atribut MCK.....	30
Tabel 4. 13 Dataset Penerima Bantuan Langsung Tunai.....	31
Tabel 4. 14 Perhitungan Dengan <i>Information Gain</i> .....	32
Tabel 4. 15 Perangkingan Atribut.....	35
Tabel 4. 16 Pemodelan <i>Naïve Bayes</i> 11 Atribut .....	35
Tabel 4. 17 Pemodelan <i>Naïve Bayes</i> 7 Atribut .....	37
Tabel 4. 18 Deskripsi Library.....	40
Tabel 5. 1 Kinerja Model <i>Naive Bayes</i> .....	51
Tabel 5. 2 <i>Classification Report</i> Klasifikasi Sebelum Seleksi Atribut.....	53
Tabel 5. 3 Skor Kepentingan .....	56
Tabel 5. 4 <i>Classification Report</i> Klasifikasi Setelah Seleksi Atribut .....	57

## DAFTAR LAMPIRAN

LAMPIRAN 1 KODE PROGRAM .....	63
LAMPIRAN 2 .....	67
LAMPIRAN 3 HASIL TURNITIN .....	68
LAMPIRAN 4 DATASET PENERIMA BANTUAN LANGSUNG TUNAI .....	<b>Error! Bookmark not defined.</b>
LAMPIRAN 5 SURAT TELAH MELAKUKAN PENELITIAN .....	71
LAMPIRAN 6 BEBAS PUSTAKA .....	72

## DAFTAR PSEUDOCODE

<i>Pseudocode 4. 1 Import Library Python .....</i>	40
<i>Pseudocode 4. 2 Menghubungkan Google Colab dengan Google Drive .....</i>	41
<i>Pseudocode 4. 3 Load Dataset.....</i>	41
<i>Pseudocode 4. 4 Konversi Dataset Excel ke CSV .....</i>	42
<i>Pseudocode 4. 5 Menampilkan beberapa baris pertama .....</i>	42
<i>Pseudocode 4. 6 Mengecek jumlah kelas .....</i>	42
<i>Pseudocode 4. 7 Mengecek jumlah Atribut Kosong.....</i>	42
<i>Pseudocode 4. 8 Mengecek jumlah atribut Pekerjaan KK.....</i>	43
<i>Pseudocode 4. 9 Konversi Data String ke Angka .....</i>	43
<i>Pseudocode 4. 10 Simpan Hasil Konversi .....</i>	43
<i>Pseudocode 4. 11 Konversi Dataset Bansos Excel ke CSV .....</i>	43
<i>Pseudocode 4. 12 Konversi Dataset Bansos Excel ke CSV .....</i>	44
<i>Pseudocode 4. 13 Setting dataset atribut dengan nilai skor yang tinggi .....</i>	44
<i>Pseudocode 4. 14 Membagi data Training dan Testing.....</i>	44
<i>Pseudocode 4. 15 Naïve Bayes untuk semua atribut .....</i>	45
<i>Pseudocode 4. 16 Confusion Matrix .....</i>	46
<i>Pseudocode 4. 17 Setting dataset 7 atribut dengan nilai skor yang tinggi .....</i>	46
<i>Pseudocode 4. 18 Membagi data Training dan Testing.....</i>	46
<i>Pseudocode 4. 19 Naïve Bayes untuk 7 atribut .....</i>	47
<i>Pseudocode 4. 20 Confusion Matrix untuk 7 atribut.....</i>	48
<i>Pseudocode 4. 21 Setting dataset 9 atribut dengan nilai skor yang tinggi .....</i>	48
<i>Pseudocode 4. 22 Membagi data Training dan Testing.....</i>	49
<i>Pseudocode 4. 23 Naïve Bayes untuk 9 atribut .....</i>	49
<i>Pseudocode 4. 24 Confusion Matrix untuk 9 atribut.....</i>	50

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pemberian bantuan sosial adalah salah satu aspek penting untuk kebijakan sosial di berbagai negara. Program-program bantuan sosial yang diberikan kepada masyarakat mempunyai dampak yang sangat signifikan dalam mencapai kesejahteraan, terutama mereka yang berada dalam kondisi ekonomi yang rentan. Kemiskinan merupakan masalah yang dihadapi negara-negara berkembang dan juga Indonesia. Masalah kemiskinan berkaitan erat dengan aspek sosial, ekonomi, pendidikan dan lainnya. Itu berarti Permasalahan kemiskinan memang merupakan permasalahan yang penting untuk dipecahkan.

Desa Muara Bone adalah salah satu desa yang berada di Kecamatan Bone, Kabupaten Bone Bolango, Provinsi Gorontalo yang menjalankan beberapa program Bantuan Sosial yang diprogramkan oleh Pemerintah Daerah untuk mengatasi kemiskinan serta meningkatkan kesejahteraan Masyarakat yang ada di daerah. Dalam proses penentuan penerima bantuan sosial Pemerintah desa yang ada di desa tersebut masih memerlukan waktu yang lama karena masih berupa lembaran-lembaran berkas hasil survey lapangan serta mengandalkan data-data dari perangkat desa yang tidak konsisten, sehingga berdampak pada penentuan penerima bantuan sosial yang belum tepat sasaran. Tentunya hal ini menjadi titik awal untuk mengklasifikasikan secara komputerisasi. Dalam konteks pengelolaan program bantuan sosial, pengklasifikasian penerima bantuan sosial merupakan suatu tugas yang penting.

Klasifikasi adalah cara mengelompokkan objek berdasarkan karakteristik objek klasifikasi. untuk melakukan klasifikasi memiliki beberapa cara baik dengan secara manual, dimana dilakukan oleh manusia tanpa bantuan dari algoritma cerdas komputer ataupun dikerjakan dengan bantuan teknologi. Klasifikasi dengan teknologi ini memiliki berbagai macam algoritma diantaranya *Naïve Bayes*, *Support Vector Machine*, *Decission Tree*, *Fuzzy* dan Jaringan Saraf Tiruan [1]. Klasifikasi ini dimaksudkan untuk mengetahui siapa yang memenuhi kriteria bantuan sosial dan siapa yang tidak. Penggunaan metode kecerdasan

buatan menjadi semakin populer untuk menyelesaikan tugas ini, banyak metode yang di pakai dalam melakukan klasifikasi, terutama metode klasifikasi menggunakan metode *Naïve Bayes*.

Algoritma *Naïve Bayes* sudah banyak digunakan oleh para peneliti sebelumnya di antaranya Diyah Utami Dkk dengan judul klasifikasi kelayakan penerima bantuan program keluarga harapan (pkh) menggunakan metode *weighted Naïve Bayes* dengan *Laplace Smoothing* dengan tingkat akurasi 95,83 % [2], Hamzah Nurriqi Fakhri Fikrillah Dkk dengan judul Klasifikasi Penerima Bansos Menggunakan Algoritma *Naïve Bayes* dengan tingkat akurasi 62,51% [3], dan Asep Surahman Dkk dengan judul implementasi algoritma *Naïve Bayes* untuk prediksi penerima bantuan sosial dengan tingkat akurasi 90,10 % [4]. dengan menggunakan metode *Naïve Bayes* memiliki kelebihan tersendiri, seperti yang sudah di sebutkan dalam penelitian Fanni Rahma Sari bahwa untuk melakukan klasifikasi menggunakan metode *Naïve Bayes* sudah terbukti mempunyai tingkat akurasi dan ketepatan yang tinggi walaupun dengan jumlah basis data yang besar [5]. Kelebihan lainnya, metode ini dalam menentukan estimasi parameter yang dibutuhkan saat proses melakukan klasifikasi sudah dapat di kerjakan cukup dengan jumlah data uji yang sedikit.

Dalam penelitian Bakti Pratama dan Astried Silvanie Akbar menyatakan metode *Naïve Bayes* memiliki kelebihan seperti, tingkat akurasi yang tinggi, sederhana, dan cepat [6]. Dari uraian penelitian yang dilakukan sebelumnya, bahwa dengan menggunakan algoritma *Naïve Bayes* sudah membuktikan bahwa metode ini dapat di gunakan dalam mengklasifikasikan penerima bantuan sosial dengan menggunakan dataset yang dikumpulkan sendiri dari Desa Muara Bone yaitu dengan atribut Pekerjaan KK, Pekerjaan Istri, Penghasilan KK, Penghasilan Istri, Jumlah Tanggungan, Sumber Listrik, Luas Tanah, Luas Bangunan, Sumber Air, MCK, Status.

Namun pada algoritma *Naïve Bayes* memiliki kelemahan dimana sangat sensitif pada saat melakukan penentuan atribut/fitur, sehingga untuk mengatasinya yaitu dengan cara dikombinasikan dengan metode seleksi fitur [7]. Sehingga salah satu tantangan utama dalam menggunakan metode klasifikasi adalah

memilih atribut atau fitur yang tepat untuk digunakan dalam proses klasifikasi. Atribut yang dipilih ketika dilakukan secara langsung maka akan memerlukan waktu yang cukup lama sehingga akan mendapatkan hasil atribut yang tidak tepat atau bahkan dapat menghilangkan atribut penting. Hal ini sudah di sebutkan secara teori dimana untuk data dengan  $n$  atribut akan terdapat sekitar  $2^n$  kemungkinan subhimpunan atribut, misalnya memiliki data dengan 10 atribut, maka harus memilih satu dari 1,024 kemungkinan subhimpunan atribut, untuk itu diperlukan teknik seleksi fitur secara otomatis [8].

Kualitas atribut yang digunakan sangat mempengaruhi kinerja model klasifikasi. Oleh karena itu, diperlukan teknik pemilihan atribut yang cermat untuk memilih atribut yang paling relevan dan informatif. Ada beberapa algoritma yang di gunakan dalam melakukan seleksi fitur seperti *Chi Square*, *N-Gram*, dan *Information Gain*. Namun pada penelitian ini akan menggunakan metode seleksi fitur *Information Gain* karena dengan metode ini dapat meningkatkan tingkat akurasi.

Hal ini sudah dibuktikan oleh Muhammad Norhalimi dan Taghfirul Azhima Yoga, dimana dalam penelitiannya hasil akurasi diperoleh dengan hanya menggunakan algoritma Naïve Bayes tanpa melakukan seleksi fitur hanya sebesar 51,73%, namun ketika dengan menggunakan metode seleksi fitur *Information Gain* meningkatkan akurasi sebesar 55,19% [9], begitu juga pada penelitian yang dilakukan oleh Amiruddin dan Rezqiwati Ishak bahwa dengan menggunakan Teknik seleksi atribut dapat meningkatkan tingkat akurasi pada beberapa algoritma *Machine Learning* (J48, *Random Forest*, MLP, SVM (SMO), dan *Naïve Bayes*), dimana hasil akurasi dari *Naïve Bayes* sebelum melakukan seleksi fitur dengan sebesar 85.67% pada scenario A dan pada scenario B sebesar 80.46%, kemudia hasil setelah dengan menggunakan teknik seleksi fitur meningkat sebesar 87.06% pada scenario A, sedangkan pada scenario B meningkat sebesar 84.13% [10].

Penggunaan *Information Gain* sebagai metode untuk melakukan seleksi atribut dalam konteks klasifikasi penerima bantuan sosial menggunakan metode *Naïve Bayes* dapat membantu meningkatkan akurasi dan efisiensi dalam



mengidentifikasi penerima yang memenuhi kriteria. Dalam konteks ini, penelitian ini bertujuan untuk mengeksplorasi penerapan *Information Gain* untuk pemilihan atribut dalam mengklasifikasikan penerima bantuan sosial menggunakan *Naïve Bayes*. Penelitian ini juga akan mengkaji dampak *Information Gain* yang diperoleh terhadap kinerja model klasifikasi dan kemampuannya untuk meningkatkan akurasi dan efisiensi administrasi program bantuan sosial.

Oleh karena itu, penelitian ini akan memberikan kontribusi yang signifikan terhadap upaya meningkatkan efektivitas dan efisiensi program bantuan sosial, serta mendorong penggunaan teknologi untuk mengambil keputusan yang lebih baik dalam konteks ini. Berdasarkan pemaparan diatas maka diperlukan untuk melakukan penelitian dengan judul “Penerapan *Information Gain* Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan *Naïve Bayes*”. Diharapkan dengan adanya penelitian ini mampu memberikan kontribusi dalam meningkatkan tingkat akurasi pada klasifikasi menggunakan *naïve bayes* dengan penerapan metode *Information Gain* sebagai algoritma untuk melakukan seleksi atribut.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang di atas maka permasalahan yang di ambil yaitu:

1. Penentuan atribut yang kemungkinan belum tentu relevan dalam melakukan klasifikasi

## **1.3 Rumusan Masalah**

Berdasarkan latar belakang, maka permasalahan dapat dirumuskan yaitu sebagai berikut:

1. Bagaimana hasil analisa seleksi atribut penting pada Penerapan *Information Gain* Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan *Naïve Bayes*?
2. Seberapa akurasi hasil klasifikasi *Naïve Bayes* apabila *Infomation Gain* diterapkan dalam seleksi atribut penting?

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis hasil seleksi atribut penting pada Penerapan *Information Gain* Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan *Naïve Bayes*.
2. Menerapkan *Information Gain* dalam seleksi atribut penting pada *Naive Bayes*, sehingga didapatkan hasil akurasi yang lebih akurat.

#### 1.5 Manfaat Penelitian

Manfaat Penelitian ini adalah sebagai berikut :

1. Manfaat Teoritis

Penelitian ini dapat menyumbangkan pengetahuan ilmiah dengan menguji efektivitas metode *Information Gain* dalam konteks klasifikasi dengan menggunakan *Naïve Bayes*. Hasil penelitian dapat membantu memperdalam pemahaman tentang peran seleksi fitur dalam pembentukan model klasifikasi.

2. Manfaat Praktis

Penelitian ini dapat membantu pemerintah atau lembaga yang terlibat dalam program bantuan sosial untuk membuat keputusan yang lebih akurat dalam penyaluran bantuan. Dengan seleksi fitur yang efektif, mereka dapat mengidentifikasi penerima bantuan dengan lebih baik.

## BAB II LANDASAN TEORI

### 2.1 Tinjauan Studi Kasus

Klasifikasi menggunakan metode *Naïve Bayes* saat ini menjadi bidang yang sudah banyak dilakukan penelitian. Dibawah ini merupakan hasil penelitian terkait sebagai referensi.

**Tabel 2. 1** Penelitian Terkait

PENELITI	JUDUL	TAHUN	HASIL
Diyah Utami, Putri Aisyiyah Rakhma Devi [2]	Klasifikasi kelayakan penerima bantuan program keluarga harapan (pkh) menggunakan metode <i>weighted naïve bayes</i> dengan <i>laplace smoothing</i>	2022	Hasil dari penelitian dengan judul klasifikasi kelayakan penerima bantuan Program Keluarga Harapan (PKH) menggunakan metode <i>weighted naïve bayes</i> dengan <i>laplace smoothing</i> dapat membantu perangkat desa dalam menentukan warga desa yang layak menerima bantuan tersebut. Metode <i>laplace smoothing</i> berhasil meminimalkan kesalahan klasifikasi, karena dapat mencegah <i>Probabilitas</i> dengan nilai 0. Dari hasil perhitungan klasifikasi menggunakan metode <i>Naïve Bayes</i> dengan <i>laplace smoothing</i> pada 56 data training dan 24 data uji didapatkan hasil pengujian evaluasi performa dengan menggunakan <i>Confusion Matrix</i> dengan nilai Accurasi 95,83%, eror 4,14%, Sensitivitas 100,00%, dan Spensifitas 94,12%. Pada penelitian selanjutnya agar hasil klasifikasi semakin bagus
Amiruddin,	Implementasi Seleksi	2022	Berdasarkan hasil penelitian

PENELITI	JUDUL	TAHUN	HASIL
Rezqiwati Ishak [8]	Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan <i>Correlation Matrix With Heatmap</i>		yang telah diuraikan di atas, maka hasil seleksi fitur menggunakan teknik <i>Correlation Matrix with Heatmap</i> pada klasifikasi waktu kelulusan Mahasiswa dari 13 atribut awal yang digunakan terseleksi menjadi 9 atribut yang memiliki relevansi atau berkontribusi terhadap hasil output yaitu : jenis kelamin, kelas, umur, SKS1, IPS1, SKS2, SKS3, SKS4 dan SKS5, sedangkan atribut input yang kurang relevan terdapat 4 atribut yaitu : IPS2, IPS3, IPS4 dan IPS5, sehingga untuk melakukan klasifikasi waktu kelulusan Mahasiswa cukup menggunakan 9 atribut yang terseleksi agar bisa mendapatkan hasil akurasi yang maksimal, khususnya di Universitas Ichsan Gorontalo. Proses klasifikasi akan dilakukan pada penelitian selanjutnya
Muhammad Norhalimi, Taghfirul Azhima Yoga Siswa[9]	Optimasi Seleksi Fitur <i>Information Gain</i> pada Algoritma <i>Naïve Bayes</i> dan <i>K-Nearest Neighbor</i>	2022	Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan yaitu Atribut keterlambatan biaya kuliah yang terbaik dipilih berdasarkan perhitungan seleksi fitur <i>Informain Gain</i> seperti fakultas, prodi, angkatan dan gender. Hasil pengujian dengan pembagian data 90:10 menggunakan algoritma <i>Naïve Bayes</i> dengan menambahkan fitur seleksi <i>Information Gain</i> mendapatkan akurasi 55,19%, sedangkan pada

PENELITI	JUDUL	TAHUN	HASIL
			<p>algoritma <i>K-Nearest Neighbor</i> dengan <i>Information Gain</i> memperoleh akurasi sebesar 50,76%. Dapat disimpulkan bahwa algoritma yang memiliki kinerja terbaik yaitu algoritma <i>Naïve Bayes</i> dan penambahan fitur seleksi <i>Information Gain</i> mampu menaikkan kinerja algoritma <i>Naïve Bayes</i> akan tetapi tidak memberi pengaruh yang signifikan terhadap kinerja algoritma <i>Naïve Bayes</i>, berbeda halnya dengan penambahan seleksi fitur <i>Information Gain</i> terhadap algoritma <i>K-Nearest Neighbor</i> yang menurunkan tingkat akurasi yang diperoleh.</p>

## 2.2 Tinjauan Pustaka

### 2.2.1 Bantuan Sosial

Pengertian bantuan sosial menurut Peraturan Menteri Keuangan Nomor 254/PMK.05/2015 tentang Belanja Bantuan Sosial pada Kementerian Negara/Lembaga adalah pengeluaran berupa transfer uang, barang atau jasa yang diberikan oleh Pemerintah kepada masyarakat miskin atau tidak mampu guna melindungi masyarakat dari kemungkinan terjadinya risiko sosial, meningkatkan kemampuan ekonomi dan/atau kesejahteraan masyarakat.

Program bansos rakyat antara lain Program Indonesia Pintar (PIP), Program Jaminan Kesehatan Nasional (JKN-KIS), Program Keluarga Harapan (PKH) dan Bansos Rastra/Bantuan Pangan Nontunai. Memperluas program bantuan pendapatan merupakan komitmen pemerintah untuk mempercepat penanggulangan kemiskinan. Hal ini tercermin dari penurunan angka kemiskinan dari 11,22% pada tahun 2015 menjadi 9,82% pada tahun 2018. Rasio Gini juga menurun dari 0,408 pada tahun 2015 menjadi 0,389 pada tahun 2018. Pada saat

yang sama, Indeks Pembangunan Manusia meningkat dari 68,90 pada tahun 2018,2014,2017 [11], jika dilihat dari data tersebut menunjukkan bahwa dengan adanya program bantuan sosial yang diberikan kepada Masyarakat miskin dapat menurunkan angka kemiskinan di Indonesia .

### 2.2.2 Data Mining

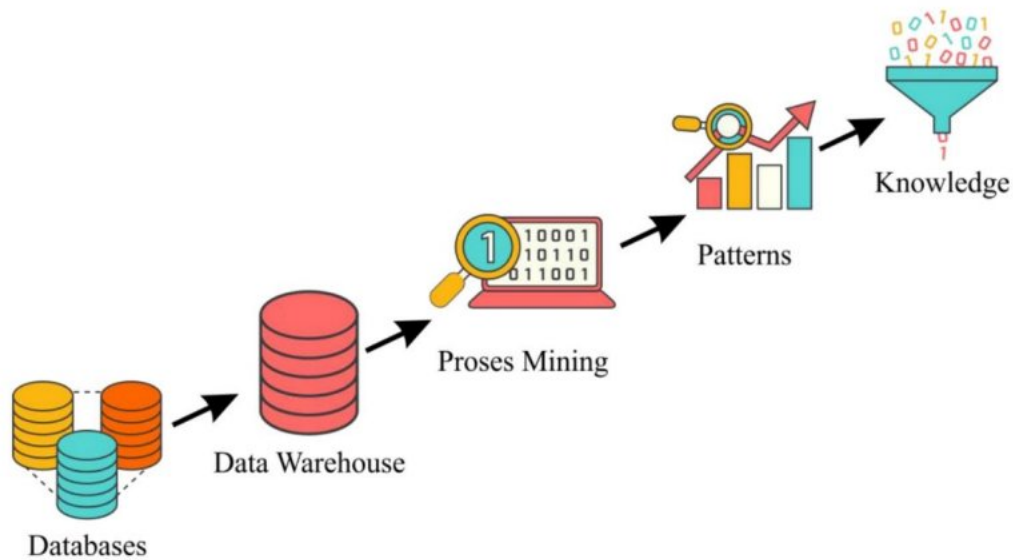
*Data Mining* adalah proses mengekstraksi informasi dan pola yang berguna dari sejumlah data yang besar. *Data Mining* meliputi pengumpulan data, ekstraksi, analisis data, dan statistik data. *Data Mining* dikenal juga dengan istilah *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain. *Data Mining* juga merupakan proses logis untuk menemukan informasi berguna. Setelah informasi dan model ditemukan, maka dapat digunakan sebagai alat bantu pengambilan keputusan dalam pengembangan bisnis.

Alat *Data Mining* dapat memberikan jawaban atas berbagai pertanyaan bisnis yang terlalu sulit untuk dipecahkan. Definisi lain dari *Data Mining* adalah proses menganalisis pola data tersembunyi dari sudut pandang berbeda dan mengklasifikasikannya menjadi data berguna yang dikumpulkan di tempat umum, gudang data untuk analisis efisien, algoritma *Data Mining*, data untuk memudahkan keputusan bisnis, dan data lainnya. . Tujuan dari *Data Mining* adalah untuk menemukan pola yang sebelumnya tidak diketahui. Setelah model tersebut diperoleh, model tersebut dapat digunakan untuk menyelesaikan berbagai permasalahan [12].

Dalam *Data Mining* di kenal juga dengan sebutan *Knowledge Discovery in Database* (KDD) adalah suatu kegiatan dimana terdapat pengumpulan data, pemakaian data, historis untuk mendapatkan keteraturan, pola bahkan hubungan dengan set data yang memiliki ukuran yang sangat besar.

Berikut Gambar Langkah-langkah saat proses *Data Mining* dimana beberapa sesuai dengan proses *Knowledge Discovery in Database* (KDD):





**Gambar 2. 1** Proses KDD

1. *Cleaning And Integration*

a. *Data Cleaning* (Pembersihan Data)

Pembersihan data merupakan proses menghilangkan noise dari data yang tidak konsisten atau bisa disebut tidak relevan. Semua data yang diperoleh dari database Perusahaan atau hasil pengujian yang tersedia tidak berisi catatan lengkap, seperti data yang hilang, data yang salah, atau sekedar kesalahan ketik. Data yang tidak relevan tersebut masih dapat di atasi dengan cara membuangnya atau yang sering disebut dengan proses cleaning. Proses pembersihan dapat mempengaruhi kinerja Teknik *Data Mining*.

b. *Data Integration* (Integrasi Data)

Integrasi data adalah proses menggabungkan data dari berbagai sumber database menjadi database yang baru. Informasi yang dibutuhkan dalam proses *Data Mining* tidak berasal dari satu database saja melainkan bisa juga berasal dari beberapa database.

2. *Selection and Transformation*

a. *Data Selection* (Seleksi Data)

Tidak semua data yang terdapat pada database digunakan karena hanya informasi relevan yang di analisis dan di ekstraksi database. Misalnya saja untuk kasus market basket analysis yang melakukan pemeriksaan factor predisposisi

klien, tidak perlu dilakukan pengambilan nama pelanggan saja, hanya dengan id pelanggan saja sudah cukup.

b. *Data Transformation* (Transformasi Data)

Transformasi data adalah proses mengubah data dan menggabungkan data ke dalam format tertentu. Sebelum diaplikasikan ke dalam data mining, membutuhkan sebuah format data khusus misalnya saja metode standar seperti analisis asosiasi dan clustering yang hanya dapat menerima data masukan yang bersifat kategoris. Oleh karena itu informasinya berupa angka-angka numerik jika mempunyai sifat kontinyu maka harus dibagi menjadi beberapa secara berkala. Proses ini biasanya sering dikenal dengan transformasi data.

3. *Proses Mining*

Pada proses ini sering dikenal dengan proses penambangan data. Proses mining adalah proses utama yang memakai metode dalam menemukan informasi berharga yang tidak terlihat atau tersembunyi dalam data.

4. *Evaluation and Presentation*

a. *Evaluasi Pola (Pattern Evaluation)*

Tugas estimasi pola adalah menemukan pola yang menarik ditemukan dalam database. Pada Langkah ini, pola karakteristik di ambil dari model klasifikasi dan evaluasi untuk menilai apakah hipotesis yang ada benar-benar tercapai. Jika terbukti hasil yang diperoleh tidak sesuai dengan hipotesis maka beberapa opsi dapat digunakan, misalnya dijadikan umpan balik untuk perbaikan proses penambangan data atau mencoba metode penambangan data lainnya.

b. *Presentasi Pengetahuan (Knowledge Presentation)*

Presentasi pengetahuan adalah visualisasi dan presentasi tentang metode yang dipakai untuk mendapatkan informasi-informasi atau data yang diambil oleh pengguna. Sebuah Langkah tahap terakhir dari proses penambangan data adalah perumusan keputusan yang diperoleh dari hasil analisis.



### 2.2.3 Naïve Bayes

Definisi *Naive Bayes* adalah klasifikasi yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yang menggunakan metode *Probabilitas* dan statistik, yaitu. memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu. *Naive Bayes* adalah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan *Probabilitas* dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data tertentu. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut tidak bergantung satu sama lain secara kondisional berdasarkan nilai awalnya.

*Probabilitas* observasi gabungan adalah produk dari *Probabilitas* individu. Keuntungan menggunakan *Naive Bayes* adalah metode ini hanya memerlukan sedikit data latih untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. *Naive Bayes* sering kali berkinerja lebih baik dari yang diharapkan dalam situasi dunia nyata yang paling kompleks. *Naive Bayes* dinilai memiliki kinerja yang sangat baik dibandingkan model klasifikasi lainnya, yaitu *Naive Bayes* memiliki akurasi yang lebih baik dibandingkan model klasifikasi lainnya [13]. Teorema Bayes adalah awal atau dasar dari metode *Naive Bayes*, dimana pada teorema bayes jika ada dua kejadian yang terpisah (misalkan X dan H), oleh karena itu Teorema Bayes dapat dirumuskan sebagai berikut:

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X	= Data dengan class yang belum diketahui
H	= Hipotesis data X merupakan suatu class spesifik
$P(H   X)$	= <i>Probabilitas</i> hipotesis H berdasarkan kondisi X
$P(H)$	= <i>Probabilitas</i> hipotesis H (Prior <i>Probabilitas</i> )
$P(X   H)$	= <i>Probabilitass</i> X berdasarkan kondisi pada hipotesis H
$P(X)$	= <i>Probabilitas</i> dari X

Mengingat berlakunya hukum *Probabilitas* total maka Teorema Bayes sering dikembangkan seperti pada rumus berikut:

$$P(H | X) = \frac{P(X|H)}{\sum_{i=1}^n P(H_i|X)} \cdot P(H) \quad (2.2)$$

Keterangan:

- i = 1,2,3,...,n jumlah data Hipotesis (prior *Probabilitas*) dimana = H1 U H2  
 U H3 ... U Hn = S  
 S = *Probabilitas* total H

Dalam penjelasan Teorema Naïve Bayes, hal yang perlu diketahui dalam proses klasifikasi akan membutuhkan beberapa petunjuk untuk mendefinisikan kelas apa yang akan cocok dengan sampel yang dianalisis tersebut. Maka dari itu, Teorema Bayes di atas disesuaikan dengan rumus sebagai berikut:

$$P(C | F_1, \dots, F_n) = \frac{P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \cdot P(C) \quad (2.3)$$

Dimana variabel C mewakili kelas, sedangkan variabel F1...Fn mewakili karakteristik instruksi yang diperlukan untuk klasifikasi. Dengan demikian, rumus tersebut menjelaskan bahwa peluang suatu sampel yang mempunyai ciri-ciri tertentu akan masuk dalam kelas C (posterior) adalah peluang munculnya kelas C (sebelum masuk sampel, sering disebut sebelumnya), dikalikan dengan peluang munculnya ciri-ciri tersebut. . sampel akan muncul di kelas C (disebut juga *likelihood*), dibagi dengan *Probabilitas* total karakteristik sampel (disebut juga *evidence*). Oleh karena itu, rumus di atas juga dapat ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (2.4)$$

Nilai pembuktiannya selalu tetap untuk setiap kelas dalam satu sampel. Nilai yang terakhir ini kemudian dibandingkan dengan nilai-nilai berikutnya dari kelas-kelas lain untuk menentukan di kelas mana sampel tersebut diklasifikasikan.

Rumus Bayes dikembangkan lebih lanjut dengan mendeskripsikan  $(C, F_1, \dots, F_n)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned}
 P(C|F_1, \dots, F_n) &= P(C).P(F_1, \dots, F_n|C) \\
 &= P(C).P(F_1|C).P(F_2, \dots, F_n|C, F_1) \\
 &= P(C).P(F_1|C).P(F_2|C, F_1).P(F_3, \dots, F_n|C, F_1, F_2) \\
 &= P(C).P(F_1|C).P(F_2|C, F_1).P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})
 \end{aligned} \tag{2.5}$$

Terlihat bahwa hasil dari perlakuan ini menyebabkan semakin kompleksnya faktor-faktor yang mempengaruhi nilai *Probabilitas*, yang hampir tidak mungkin untuk dianalisis secara individual. Oleh karena itu perhitungan ini sulit dilakukan. Asumsi independensi yang sangat tinggi (naif) digunakan di sini, bahwa setiap ordo  $(F_1, F_2, \dots, F_n)$  bebas (independen) satu sama lain. Berdasarkan asumsi-asumsi ini persamaan berikut berlaku:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i).P(F_j)}{P(F_j)} = P(F_i) \tag{2.6}$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C) \tag{2.7}$$

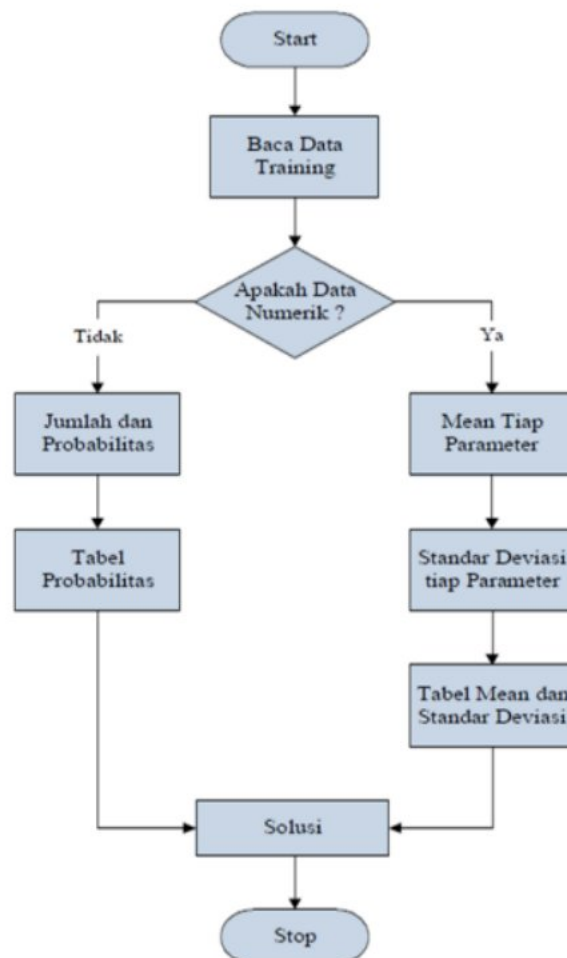
Atau juga dapat ditulis dengan notasi

$$P(C|F_1, F_2, F_3, \dots, F_n) = P(C) \prod_{i=1}^n P(F_i|C) \tag{2.8}$$

Dimana penjabarannya sebagai berikut:

$$\begin{aligned}
 P(c|F) \\
 &= P(f_1|c).P(f_2|c).P(f_3|c) \dots P(f_n|c).P(c)
 \end{aligned} \tag{2.9}$$

Persamaan di atas adalah suatu model dari algoritma Teorema Naïve Bayes yang kemudian akan di pakai pada proses klasifikasi. Dalam metode Naïve Bayes juga memiliki alur yang dapat dilihat pada Gambar 2.1 sebagai berikut :



**Gambar 2. 2** Alur Metode *Naive Bayes*

Penjelasan berdasarkan gambar di atas adalah sebagai berikut:

1. Baca data training
2. Hitung jumlah dan *Probabilitas*, tetapi jika data numerik maka:
  - a. Carilah mean dan standar deviasi dari setiap parameter yang berupa data numerik.

Adapun persamaan yang digunakan untuk menghitung nilai rata-rata (mean) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (2.10)$$

Atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2.11)$$

Keterangan:

$\mu$  = rata-rata hitung (mean)

$x_i$  = nilai sampel ke-1

$n$  = jumlah sampel

Dan persamaan dalam menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (2.12)$$

Keterangan:

$\sigma$  = standar deviasi

$X_i$  = nilai x ke-i

$\mu$  = rata-rata hitung

$n$  = jumlah sampel

- b. Untuk mencari nilai *Probabilitas*, hitunglah jumlah data yang cocok pada kelas yang sama dibagi dengan jumlah data pada kelas tersebut.
3. Mendapatkan nilai dalam tabel mean, standar deviasi dan *Probabilitas*
4. Solusi yang dihasilkan.

### 2.2.4 Confusion Matrix

*Confusion Matrix* adalah matriks dua dimensi yang menjelaskan perbandingan hasil prediksi dengan kenyataan. juga menegaskan bahwa *Confusion Matrix* adalah spreadsheet yang digunakan untuk memperkirakan model klasifikasi item yang mana diprediksi dengan benar dan salah. Evaluasinya menggunakan *Confusion Matrix* untuk tingkat akurasi kinerja model klasifikasi. untuk Pengukuran kinerja model klasifikasi dilakukan perbandingan dengan semua data tes yang diklasifikasikan valid untuk sejumlah besar data uji. Akurasi adalah ukuran rasio prediksi yang benar terhadap total sampel dievaluasi. Dengan kata lain, akurasi adalah derajat kedekatan nilai prediksi dengan nilai sebenarnya [14]. Ada beberapa cara untuk mengukur kinerja dalam *Data Mining* salah satunya dengan model yang di hasilkan menggunakan Confusion Matriks (akurasi). Confusion matriks merupakan metode yang digunakan untuk membuat perhitungan yang tepat untuk konsep *Data Mining* [15]. Dalam metode ini memiliki empat istilah untuk merepresentasikan hasil saat proses klasifikasi sebagai berikut:

**Tabel 2. 2** *Confusion Matrix*

Kelas	Klasifikasi Positif	Klasifikasi Negatif
Positif	TP (True Positif)	FN (False Negatif)
Negatif	FP (False Positif)	TN (True Negatif)

Keterangan:

1. *True Negatif* (TN) adalah sejumlah data yang sudah didekteksi oleh sistem dengan benar.
2. *False Positif* (FP) adalah suatu data negative tetapi sudah dideteksi pada sistem sebagai data *positif*
3. *True Positif* (TP) adalah suatu data positif yang dideteksi oleh sistem dengan benar.
4. *False Negatif* (FN) adalah data *positif* tetapi dideteksi sebagai data *negative*.

Berdasarkan tabel 2.2 berikut perhitungan Akurasi, *Precision*, *Recall* dan *F-Measure* yang sesuai dengan tabel di atas.

1. Pada nilai akurasi ini akan mendeksripsikan ke akuratan dari suatu sistem yang dapat mengklasifikasikan data dengan benar. Nilai akurasi adalah membandingkan antara data yang terklasifikasikan benar dengan seluruh data yang ada. Maka persamaannya sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (2.13)$$

2. Untuk nilai Presisi akan menggambarkan sejumlah data pada data yang berkategori positif dan diklasifikasikan dengan benar kemudian akan dibagi dengan total data pada klasifikasi positif. Sehingga persamaannya sebagai berikut.

$$Precision = \frac{TP}{FP + TP} * 100\% \quad (2.14)$$

3. *Recall* atau nilai *Sensitivity*, dimana pada nilai ini akan menunjukan jumlah persen dari data yang berkategori *positif* yang diklasifikasikan dengan benar oleh sistem. Maka persamaannya adalah sebagai berikut.

$$Recall = \frac{TP}{TP + FN} * 100\% \quad (2.15)$$

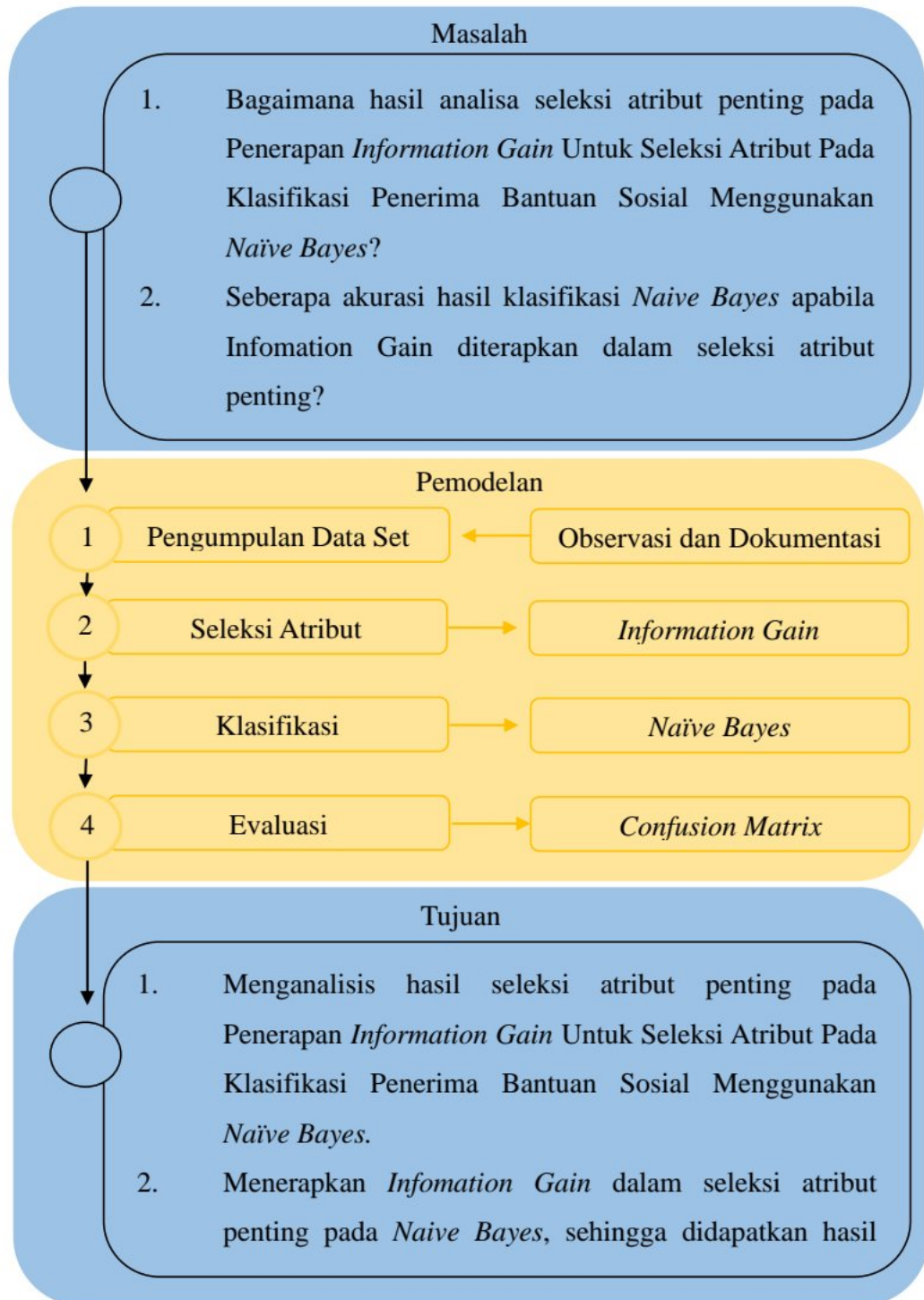
4. Sedangkan untuk *F-Measure* merupakan suatu perhitungan untuk nilai pada keterangan *retrival* (temu Kembali) dimana akan mengkombinasikan

dengan nilai *recall* dan presisi. Sehingga berikut rumus untuk menghitung *F-Measure*.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2.16)$$



### 2.3 Kerangka Pikir



**Gambar 2. 3** Kerangka Pikir

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian**

Dilihat dari tingkat penerapannya, maka penelitian ini adalah penelitian terapan. Jika dipandang dari jenis informasi yang diolah, maka penelitian ini bersifat Kuantitatif. Kemudian jika dilihat dari pengolahan datanya, maka penelitian ini merupakan penelitian konfirmatori.

Penelitian ini menggunakan metode penelitian Studi kasus. Sehingga jenis penelitian ini bersifat deskriptif. Subjek penelitian ini adalah data penduduk penerima bantuan sosial di Desa Muara Bone, Kec. Bone, Provinsi Gorontalo. Sedangkan untuk objek pada penelitian ini adalah penentuan atribut yang paling berpengaruh dalam menentukan kelayakan penerima bantuan sosial. Penelitian ini dimulai dari bulan September 2023 sampai dengan selesai, dimana penelitian ini berlokasi di Jln. Trans Sulawesi, Desa Muara Bone, Kec. Bone, Kab. Bone Bolango, Provinsi Gorontalo.

#### **3.2 Pengumpulan Data**

Sumber data yang akan digunakan untuk penelitian ini adalah:

##### **1. Data Primer**

Data primer di peroleh lewat wawancara serta observasi langsung di lapangan. Data primer ini adalah informasi yang dikumpulkan peneliti langsung dari sumbernya tanpa perantara, atau langsung melalui responden. Data ini berupa data penduduk penerima bantuan sosial dari tahun 2022 sampai 2023.

Adapun atribut atau variabel beserta tipe datanya masing-masing dapat dilihat pada tabel berikut:

**Tabel 3. 1** Atribut Data Penerima Bantuan Langsung Tunai

NO	Name	Value	Keterangan
1	Pekerjaan KK	Tidak Bekerja, Petani, Nelayan, Peg. Swasta, Wirausaha, Lainnya	Variabel Input
2	Pekerjaan Istri	Tidak Bekerja, Petani, Wirausaha, Lainnya	Variabel Input
3	Penghasilan KK	Tidak Berpenghasilan, < Rp.	Variabel

NO	Name	Value	Keterangan
		250.000, Rp. 250.001 - Rp. 500.000, Rp. 500.001 - Rp. 750.000, Rp. 750.001 - Rp. 1.000.000, >Rp. 1.000.000	Input
4	Penghasilan Istri	Tidak Berpenghasilan, < Rp. 250.000, Rp. 250.001 - Rp. 500.000, Rp. 500.001 - Rp. 750.000, Rp. 750.001 - Rp. 1.000.000, >Rp. 1.000.000	Variabel Input
5	Jumlah Tanggungan	1-2 Orang, 3-4 Orang, >4	Variabel Input
6	Kepemilikan Rumah	Sendiri, menumpang, Tidak memiliki	Variabel Input
7	Sumber Listrik	Tidak ada, Menumpang Tetangga, PLN, PLN & Genset, Tenaga Surya	Variabel Input
8	Luas Tanah	<25 M2, 25-50 M2, 50-99 M2, 100-200 M2, >200 M2	Variabel Input
9	Luas Bangunan	<25 M2, 25-50 M2, 50-99 M2, 100-200 M2, >200 M2	Variabel Input
10	Sumber air	Sungai/mata air, Sumur, PDAM, Kemasan	Variabel Input
11	MCK	Berbagi Pakai, Kepemilikan Sendiri Diluar, Kepemilikan Sendiri Didalam	Variabel Input
12	Status	Terima, Tidak Terima	Variabel Output

## 2. Data sekunder

Untuk data sekunder merupakan data yang terlebih dahulu diolah kemudian diperoleh informasi tambahan, sedangkan sumber data sekunder diperoleh melalui dokumentasi dan studi kepustakaan, membaca berbagai referensi dari para ahli, jurnal, website serta berbagai sumber lain yang membahas tentang metode klasifikasi dan seleksi atribut.

Adapun Teknik yang akan digunakan dalam pengumpulan data sekunder sebagai berikut:

### 3.2.1 Observasi

Observasi adalah suatu metode pengumpulan data dimana lokasi penelitian akan diamati atau ditinjau secara langsung untuk membantu peneliti menyelesaikan suatu penelitian yang sedang berlangsung. Tentunya dalam proses ini akan berdasarkan jurnal serta referensi terkait. Penentuan kelayakan penerima bantuan sosial di Desa Muara Bone, berdasarkan temuan peneliti, memerlukan waktu yang cukup lama karena masih mengandalkan lembaran-lembaran kertas hasil survey lapangan dan data dari perangkat desa yang belum konsisten, serta atribut data yang dipakai masih banyak kesamaan datanya sehingga belum efektif dalam menentukan kelayakan penerima bantuan sosial.

### 3.2.2 Wawancara

Wawancara adalah suatu Teknik untuk mengumpulkan data melalui proses interaksi antara peneliti dan pihak responden, pada umumnya proses ini dilakukan dengan tanya jawab. Dalam proses pengumpulan data dengan Teknik wawancara, peneliti mengambil dari beberapa penelitian terkait sehingga dengan wawancara ini peneliti dapat menggali informasi atau data, serta keterangan dari subjek penelitian.

### 3.2.3 Dokumentasi

Pada proses dokumentasi ini peneliti masih menggunakan jurnal-jurnal penelitian terkait, literature, atau referensi agar peneliti bisa mendapatkan data yang berhubungan dengan subjek dan objek yang sedang diteliti. Berdasarkan penelitian terkait sebelumnya bahwa data yang diambil atau dikumpulkan bersumber dari dokumen yang berkaitan perihal segala yang ada pada permasalahan yang sedang diteliti.

## 3.3 Pengembangan Model

Pada penelitian ini tahap-tahap atau prosedur dalam klasifikasi menggunakan metode Naïve bayes untuk mengklasifikasikan kelayakan penerima Bantuan Langsung Tunai di Desa Muara Bone. Kemudian untuk mendukung serta meningkatkan tingkat akurasi akan digunakan metode *Information Gain* sebagai algoritma untuk melakukan seleksi atribut sehingga tujuannya untuk

menentukan atribut mana saja yang paling berpengaruh untuk digunakan dalam proses klasifikasi dan dengan atribut tersebut dapat meningkatkan tingkat akurasi. Dalam hal ini juga peneliti akan menggunakan tools Python sebagai software pendukung dalam proses klasifikasi.

### **3.4 Evaluasi Model**

Pada evaluasi model ini merupakan suatu model yang telah dihasilkan lalu akan dievaluasi menggunakan *Confusion Matrix* sebagai metode untuk mengetahui tingkat akurasi, *recall* serta *precesion*.

### **3.5 Analisis Deskriptif**

Analisis ini adalah metode untuk analisis statistik dimana tujuannya agar supaya mampu memberikan suatu gambaran tentang subjek penelitian berdasarkan data variabel yang diperoleh dari suatu subjek tertentu.



## BAB IV HASIL PENELITIAN

### 4.1 Hasil Pengumpulan Data

Berikut Dataset Penerima Bantuan Langsung Tunai antara lain, dimana data yang diperoleh dari Kantor Desa Muara Bone Sebanyak 134 record dalam bentuk file excel. Bentuk dataset ditunjukkan pada Tabel 4.1

**Tabel 4. 1** Hasil Pengumpulan Data Set

No	Nama KK	Pekerjaan KK	Pekerjaan Istri	Penghasilan KK	Penghasilan Istri	Jumlah Tanggungan
1.	Alimin Dunggio	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	3-4 Orang
2.	M. TAKDIR	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	>4 Orang
3.	Erwin S.M.Mosey	Petani	Tidak Bekerja	Rp. 500.001 - Rp. 750.000	Tidak Berpenghasilan	3-4 Orang
4.	ALPIAN POTABUGA	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	1-2 Orang
5.	Zainal Abidin	Petani	Lainnya	Rp. 250.001 - Rp. 500.000	>Rp. 1.000.000	3-4 Orang
6.	Anis solong	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	3-4 Orang
7.	Herman kadir	Lainnya	Tidak Bekerja	Rp. 750.001 - Rp. 1.000.000	Tidak Berpenghasilan	1-2 Orang
8.	Bodiono F. Nurkamen	Petani	Lainnya	Rp. 250.001 - Rp. 500.000	Rp. 250.001 - Rp. 500.000	3-4 Orang
9.	Ilwan Tangkudung	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan	1-2 Orang
10.	Abdul Haris Yusuf	Petani	Tidak Bekerja	Rp. 500.001 - Rp. 750.000	Tidak Berpenghasilan	3-4 Orang
.....	.....	.....	.....	.....	.....	.....
134.	Ferri Setyawan Laiya	Tidak Bekerja	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Rp. 250.001 - Rp. 500.000	1-2 Orang

No	Kepemilikan Rumah	Sumber Listrik	Luas Tanah	Luas Bangunan	Sumber Air	MCK	Status
1.	Sendiri	PLN	< 25 M2	< 25 M2	PDAM	Kepemilikan Sendiri Didalam	Terima
2.	Sendiri	PLN	100-200 M2	25-50 M2	Sumur	Kepemilikan Sendiri Diluar	Terima
3.	Sendiri	PLN	25-50 M2	< 25 M2	PDAM	Kepemilikan Sendiri Didalam	Terima
4.	Sendiri	PLN	< 25 M2	< 25 M2	SUmur	Kepemilikan Sendiri Diluar	Terima
5.	Sendiri	PLN	100-200 M2	50-99 M2	Sumur	Kepemilikan Sendiri Didalam	Tidak Terima
6.	Sendiri	PLN	>200 M2	50-99 M2	Sumur	Kepemilikan Sendiri Diluar	Terima
7.	Sendiri	Tidak Ada	>200 M2	25-50 M2	Sumur	Kepemilikan Sendiri Didalam	Tidak Terima
8.	sendiri	PLN	< 25 M2	< 25 M2	Sumur	Kepemilikan Sendiri Diluar	Terima
9.	Sendiri	PLN	< 25 M2	< 25 M2	Sumur	Kepemilikan Sendiri Diluar	Terima
10.	Sendiri	PLN	>200 M2	>200 M2	Sumur	Kepemilikan Sendiri Didalam	Terima
.....	.....	.....	.....	.....	.....	.....	.....
134.	Menumpang	PLN	25-50 M2	25-50 M2	PDAM	Kepemilikan Sendiri Didalam	Terima

## 4.2 Hasil Pemodelan

Ketika melakukan pemodelan menggunakan metode *Naive Bayes* maka terlebih dahulu dilakukan beberapa *Preprocessing* data terhadap dataset pada tabel 4.1 diatas khususnya pada atribut dataset awal penerima bantuan langsung tunai dilakukan konversi data dari jenis kategorikal menjadi numerik sehingga data akan menjadi satu jenis type yaitu numerik.



#### 4.2.1 Data Preprocessing

Data Preprocessing dilakukan untuk memastikan kualitas data yang digunakan dalam analisis dan pembuatan model. Tujuan utama dari data preprocessing adalah untuk mempersiapkan data agar siap digunakan dalam proses selanjutnya, serta untuk meningkatkan akurasi dan efisiensi model klasifikasi. Langkah-langkah yang diambil dalam preprocessing data adalah sebagai berikut:

##### 1. Konversi Data pada Atribut Pekerjaan KK

Pada data atribut Pekerjaan KK untuk konversi nilainya ditunjukkan pada Tabel 4.2 berikut :

**Tabel 4. 2** Konversi Atribut Pekerjaan KK

Pekerjaan KK	Nilai
Tidak Bekerja	1
Petani	2
Nelayan	3
Peg. Swasta	4
Wirausaha	5
Lainnya	6

##### 2. Konversi Data pada Atribut Pekerjaan Istri

Pada data atribut Pekerjaan Istri untuk konversi nilainya ditunjukkan pada Tabel 4.3 berikut :

**Tabel 4. 3** Konversi Atribut Pekerjaan Istri

Pekerjaan Istri	Nilai
Tidak Bekerja	1
Petani	2
Wirausaha	3
Lainnya	4

##### 3. Konversi Data pada Atribut Penghasilan KK

Pada data atribut Penghasilan KK untuk konversi nilainya ditunjukkan pada Tabel 4.4 berikut :

**Tabel 4. 4** Konversi Atribut Penghasilan KK

Penghasilan KK	Nilai
Tidak Berpenghasilan	1
< Rp. 250.000	2
Rp. 250.001 - Rp. 500.000	3
Rp. 500.001 - Rp. 750.000	4
Rp. 750.001 - Rp. 1.000.000	5
>Rp. 1.000.000	6

#### 4. Konversi Data pada Atribut Penghasilan Istri

Pada data atribut Penghasilan Istri untuk konversi nilainya ditunjukkan pada Tabel 4.5 berikut :

**Tabel 4. 5** Konversi Atribut Penghasilan Istri

Penghasilan Istri	Nilai
Tidak Berpenghasilan	1
< Rp. 250.000	2
Rp. 250.001 - Rp. 500.000	3
Rp. 500.001 - Rp. 750.000	4
Rp. 750.001 - Rp. 1.000.000	5
>Rp. 1.000.000	6

#### 5. Konversi Data pada Atribut Jumlah Tanggungan

Pada data atribut Jumlah Tanggungan untuk konversi nilainya ditunjukkan pada Tabel 4.6 berikut :

**Tabel 4. 6** Konversi Atribut Jumlah Tanggungan

Jumlah Tanggungan	Nilai
1-2 Orang	1
3-4 Orang	2
>4 Orang	3

#### 6. Konversi Data pada Atribut Luas Tanah

Pada data atribut Luas Tanah untuk konversi nilainya ditunjukkan pada Tabel 4.7 berikut :

**Tabel 4. 7** Konversi Atribut Luas Tanah

Luas Tanah	Nilai
< 25 M2	1
25-50 M2	2
50-99 M2	3
100-200 M2	4
>200 M2	5

### 7. Konversi Data pada Atribut Luas Bangunan

Pada data atribut Luas Bangunan untuk konversi nilainya ditunjukkan pada Tabel 4.8 berikut :

**Tabel 4. 8** Konversi Atribut Luas Bangunan

Luas Bangunan	Nilai
< 25 M2	1
25-50 M2	2
50-99 M2	3
100-200 M2	4
>200 M2	5

### 8. Konversi Data pada Atribut Kepemilikan Rumah

Pada data atribut Kepemilikan Rumah untuk konversi nilainya ditunjukkan pada Tabel 4.8 berikut :

**Tabel 4. 9** Konversi Atribut Kepemilikan Rumah

Kepemilikan Rumah	Nilai
Tidak Memiliki	1
Menumpang	2
Sendiri	3

### 9. Konversi Data pada Atribut Sumber Listrik

Pada data atribut Sumber Listrik untuk konversi nilainya ditunjukkan pada Tabel 4.8 berikut :

**Tabel 4. 10** Konversi Atribut Sumber Listrik

Sumber Listrik	Nilai
Tidak Ada	1
Menumpang Tetangga	2
PLN	3
PLN dan Genset	4
Tenaga Surya	5

### 11. Konversi Data pada Atribut Sumber Air

Pada data atribut Sumber Air untuk konversi nilainya ditunjukkan pada Tabel 4.8 berikut :

**Tabel 4. 11** Konversi Atribut Sumber Air

Sumber Air	Nilai
Sungai/Mata Air	1
Sumur	2
PDAM	3
Kemasan	4

### 11. Konversi Data pada Atribut MCK

Pada data atribut MCK untuk konversi nilainya ditunjukkan pada Tabel 4.8 berikut :

**Tabel 4. 12** Konversi Atribut MCK

MCK	Nilai
Berbagi Pakai	1
Kepemilikan Sendiri Diluar	2
Kepemilikan Sendiri Didalam	3

Setelah dilakukan konversi untuk semua atribut yang bernilai kategorikal, maka hasil akhir dataset yang akan digunakan ditunjukan pada Tabel 4.13.

**Tabel 4. 13** Dataset Penerima Bantuan Langsung Tunai

No	Nama KK	Pekerjaan KK	Pekerjaan Istri	Penghasilan KK	Penghasilan Istri	Jumlah Tanggungan
1.	Alimin Dunggio	2	1	3	1	2
2.	M. TAKDIR	2	1	3	1	1
3.	Erwin S.M.Mos ey	2	1	4	1	2
4.	ALPIAN POTAB UGA	2	1	3	1	3
5.	Zainal Abidin	2	4	3	6	2
6.	Anis solong	2	1	3	1	2
7.	Herman kadir	6	1	5	1	3
8.	Bodiono F. Nurkami den	2	4	3	3	2
9.	Ilwan	2	1	3	1	3
10.	Abdul Haris Yusuf	2	1	4	1	2
.....	.....	.....	.....	.....	.....	.....
134.	Ferri Setyawa n Laiya	1	1	2	2	3

No	Kepemilikan Rumah	Sumber Listrik	Luas Tanah	Luas Bangunan	Sumber Air	MCK	Status
1.	3	3	1	1	3	3	Terima
2.	3	3	4	2	2	2	Terima
3.	3	3	2	1	3	3	Terima
4.	3	3	1	1	2	2	Terima
5.	3	3	4	3	2	3	Tidak Terima
6.	3	3	5	3	2	2	Terima
7.	3	1	5	2	2	3	Tidak Terima
8.	3	3	1	1	2	2	Terima
9.	3	3	1	1	2	2	Terima
10.	3	3	1	1	3	3	Terima
.....	.....	.....	.....	.....	.....	.....	.....
134.	2	3	2	2	3	3	Terima

#### 4.2.2 Pemodelan dengan Metode *Information Gain*

Sebelum dilakukan Klasifikasi dengan Metode *Naive Bayes* sebelumnya dilakukan seleksi Atribut yang penting dengan menggunakan Metode *Information Gain*. Adapun Dataset yang digunakan sebagai sampel dapat dilihat pada Tabel 4.13. Adapun hasil perhitungan dengan metode *Information Gain* diantaranya sebagai berikut :

##### 1. Perhitungan Metode *Information Gain* pada Semua Atribut

Pada tahap ini akan Menghitung proporsi masing-masing kelas Terima dan Tidak Terima serta menentukan *Information Gain*nya. Berikut perhitungan yang dilakukan pada aplikasi *Microsoft Excel* pada Tabel 4.14 dibawah ini.

**Tabel 4. 14** Perhitungan Dengan *Information Gain*

Atribut		Terima	Tidak Terima	Jumlah	Entropy	<i>Information Gain</i>
Pekerjaan KK	Tidak Bekerja	6	4	10	0.97	0.38
	Petani	45	3	48	0.34	
	Nelayan	5	0	5	0.00	
	Peg. Swasta	1	2	3	0.92	
	Wirausaha	1	5	6	0.65	
	Lainnya	9	26	35	0.82	
Total		67	40	107	0.95	
Pekerjaan Istri	Tidak Bekerja	52	26	78	0.92	0.06
	Petani	6	2	8	0.81	
	Wirausaha	0	3	3	0.00	
	Lainnya	9	9	18	1.00	
Total		67	40	107	0.95	
Penghasilan KK	Tidak Berpenghasilan	5	4	9	0.99	0.21
	'< Rp. 250.000	21	1	22	0.27	
	Rp. 250.001 - Rp. 500.000	28	9	37	0.80	
	Rp. 500.001 - Rp. 750.000	8	13	21	0.96	
	Rp. 750.001 - Rp. 1.000.000	1	5	6	0.65	
	'>Rp. 1.000.000	4	8	12	0.92	
Total		67	40	107	0.95	
Penghasilan Istri	Tidak Berpenghasilan	33	24	57	0.98	0.21



Atribut		Terima	Tidak Terima	Jumlah	Entropy	Information Gain
	'<Rp. 250.000	27	3	30	0.47	
	Rp. 250.001 - Rp.500.000	6	2	8	0.81	
	Rp. 500.001 - Rp.750.000	1	3	4	0.81	
	Rp. 750.001 - Rp. 1.000.000	0	2	2	0.00	
	'>Rp. 1.000.000	0	6	6	0.00	
Total		67	40	107	0.95	
Jumlah Tanggungan	1-2 Orang	28	12	40	0.88	0.03
	3-4 Orang	25	12	37	0.91	
	>4 Orang	14	16	30	1.00	
Total		67	40	107	0.95	
Kepemilikan Rumah	Tidak Memiliki	2	0	2	0.00	0.02
	Menumpang	7	7	14	1.00	
	Sendiri	58	33	91	0.94	
Total		67	40	107	0.95	
Sumber Listrik	Tidak Ada	0	1	1	0.00	0.04
	Menumpang tetangga	1	0	1	0.00	
	PLN	63	39	102	0.96	
	PLN dan Genset	3	0	3	0.00	
	Tenaga Surya	0	0	0	0.00	
Total		67	40	107	0.95	
Luas Tanah	'< 25 M2	32	5	37	0.57	0.37
	25-50 M2	5	4	9	0.99	
	50-99 M2	8	3	11	0.85	
	100-200 M2	10	13	23	0.99	
	'>200 M2	12	15	27	0.99	
Total		67	40	107	0.95	
Luas Bangunan	'< 25 M2	38	6	55	0.72	0.16
	25-50 M2	14	9	23	0.97	
	50-99 M2	4	12	16	0.81	
	100-200 M2	5	5	10	1.00	
	'>200 M2	6	8	17	1.04	
Total		67	40	107	0.95	
Sumber Air	Sungai/Mata Air	4	3	7	0.99	0.06
	Sumur	46	18	64	0.86	
	PDAM	15	12	27	0.99	
	Kemasan	2	7	9	0.76	
Total		67	40	107	0.95	



Atribut		Terima	Tidak Terima	Jumlah	Entropy	Information Gain
MCK	Berbagi Pakai	5	3	8	0.95	0.07
	Kepemilikan Sendiri Diluar	25	4	29	0.58	
	Kepemilikan Sendiri Didalam	37	33	70	1.00	
Total		67	40	107	0.95	

Adapun sebagian penjelasan terkait perhitungan dari tabel diatas sebagai berikut :

1. Menghitung Entropy Atribut Pekerjaan KK pada Sub Atribut Tidak Bekerja

Terima : 6 dari 10 Sampel

Tidak Terima : 4 dari 10 Sampel

Formula Entropy :  $Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i$

$$Entropy(S) = \left( -\frac{6}{10} \log_2 \frac{6}{10} \right) + \left( -\frac{4}{10} \log_2 \frac{4}{10} \right)$$

Perhitungan :

$$Entropy(S) = (-0.6 \log_2 0.6) + (-0.4 \log_2 0.4)$$

$$Entropy(S) = (-0.6 * -0.74) + (-0.4 * -1.33)$$

$$Entropy(S) = (0.444 + 0.532)$$

$$Entropy(S) = 0.97$$

2. Setelah menghitung Entropy maka Langkah selanjutnya adalah menghitung *Information Gain*nya dengan perhitungan sebagai berikut.

Formula *Information Gain*:

$$IG(S,A) = Entropy(S) - \sum_{Value(A)} \left( \frac{S_v}{S} \right) Entropy(S_v)$$

$$\begin{aligned} IG(S, Pekerjaan KK) &= 0.95 - \left( \left( \frac{10}{107} \right) * 0.97 \right) + \left( \left( \frac{48}{107} \right) * 0.34 \right) \\ &\quad + \left( \left( \frac{5}{107} \right) * 0 \right) + \left( \left( \frac{3}{107} \right) * 0.92 \right) + \left( \left( \frac{6}{107} \right) * 0.65 \right) \\ &\quad + \left( \left( \frac{35}{107} \right) * 0.82 \right) \end{aligned}$$

$$IG(S, Pekerjaan KK) = 0.95 - (0.09 + 0.15 + 0 + 0.02 + 0.04 + 0.27)$$

$$IG(S, Pekerjaan KK) = 0.95 - 0.57$$

$$IG(S, Pekerjaan KK) = 0.38$$

3. Setelah dilakukan perhitungan *Information Gain* pada semua Atribut hasilnya diurutkan mulai dari nilai IG nya tertinggi sampai yang terendah yang ditunjukkan pada Tabel 4.15.

**Tabel 4. 15** Perangkingan Atribut

No	Atribut	<i>Information Gain</i>
1.	Pekerjaan KK	0.38
2.	Luas Tanah	0.37
3.	Penghasilan KK	0.21
4.	Penghasilan Istri	0.21
5.	Luas Bangunan	0.16
6.	MCK	0.07
7.	Pekerjaan Istri	0.06
8.	Sumber Air	0.00
9.	Sumber Listrik	0.00
10.	Jumlah Tanggungan	0.00
11.	Kepemilikan Rumah	0.00

#### 4.2.3 Pemodelan dengan *Naïve Bayes*

Pemodelan menggunakan Metode *Naïve Bayes* dilakukan penentuan atribut yang ingin dipakai dengan memilih atribut yang sudah di seleksi menggunakan Metode *Information Gain*. Adapun akan dilakukan beberapa uji coba menggunakan 11 atribut dan 7 Atribut berdasarkan nilai gain tertinggi. Berikut perhitungan untuk 11 atribut dapat dilihat pada Tabel 4.16.

**Tabel 4. 16** Pemodelan *Naïve Bayes* 11 Atribut

Atribut		Terima	Tidak Terima
Pekerjaan KK	Tidak Bekerja	0.09	0.10
	Petani	0.67	0.08
	Nelayan	0.07	0.00
	Peg. Swasta	0.01	0.05
	Wirausaha	0.01	0.13
	Lainnya	0.13	0.65

Atirbut		Terima	Tidak Terima
Luas Tanah	'< 25 M2	0.48	0.13
	25-50 M2	0.07	0.10
	50-99 M2	0.12	0.08
	100-200 M2	0.15	0.33
	'>200 M2	0.18	0.38
Penghasilan KK	Tidak Berpenghasilan	0.07	0.10
	'< Rp. 250.000	0.31	0.03
	Rp. 250.001 - Rp. 500.000	0.42	0.23
	Rp. 500.001 - Rp. 750.000	0.12	0.33
	Rp. 750.001 - Rp. 1.000.000	0.01	0.13
	'>Rp. 1.000.000	0.06	0.20
Penghasilan Istri	Tidak Berpenghasilan	0.49	0.60
	'<Rp. 250.000	0.40	0.08
	Rp. 250.001 - Rp.500.000	0.09	0.05
	Rp. 500.001 - Rp.750.000	0.01	0.08
	Rp. 750.001 - Rp. 1.000.000	0.00	0.05
	'>Rp. 1.000.000	0.00	0.15
Luas Bangunan	'< 25 M2	0.57	0.15
	25-50 M2	0.21	0.23
	50-99 M2	0.06	0.30
	100-200 M2	0.07	0.13
	'>200 M2	0.09	0.20
MCK	Berbagi Pakai	0.07	0.08
	Kepemilikan Sendiri Diluar	0.37	0.10
	Kepemilikan Sendiri Didalam	0.55	0.83
Pekerjaan Istri	Tidak Bekerja	0.78	0.65
	Petani	0.09	0.05
	Nelayan	0.00	0.00
	Peg. Swasta	0.00	0.00
	Wirausaha	0.00	0.08
	Lainnya	0.13	0.23
Sumber Air	Sungai/Mata Air	0.06	0.08
	Sumur	0.69	0.45
	PDAM	0.22	0.30
	Kemasan	0.03	0.18
Sumber Listrik	Tidak Ada	0.00	0.03
	Menumpang tetangga	0.01	0.00
	PLN	0.94	0.98
	PLN dan Genset	0.04	0.00
	Tenaga Surya	0.00	0.00
Jumlah Tanggungan	1-2 Orang	0.41791	0.30

Atirbut		Terima	Tidak Terima
	3-4 Orang	0.373134	0.30
	>4 Orang	0	0.00
Kepemilikan Rumah	Tidak Memiliki	0.029851	0.00
	Menumpang	0.104478	0.18
	Sendiri	0.865672	0.83

Adapun perhitungan untuk 7 atribut dapat dilihat pada tabel 4.17 berikut:

**Tabel 4. 17** Pemodelan *Naïve Bayes* 7 Atribut

Atirbut		Terima	Tidak Terima
Pekerjaan KK	Tidak Bekerja	0.09	0.10
	Petani	0.67	0.08
	Nelayan	0.07	0.00
	Peg. Swasta	0.01	0.05
	Wirausaha	0.01	0.13
	Lainnya	0.13	0.65
Luas Tanah	'< 25 M2	0.48	0.13
	25-50 M2	0.07	0.10
	50-99 M2	0.12	0.08
	100-200 M2	0.15	0.33
	'>200 M2	0.18	0.38
Penghasilan KK	Tidak Berpenghasilan	0.07	0.10
	'< Rp. 250.000	0.31	0.03
	Rp. 250.001 - Rp. 500.000	0.42	0.23
	Rp. 500.001 - Rp. 750.000	0.12	0.33
	Rp. 750.001 - Rp. 1.000.000	0.01	0.13
	'>Rp. 1.000.000	0.06	0.20
Penghasilan Istri	Tidak Berpenghasilan	0.49	0.60
	'<Rp. 250.000	0.40	0.08
	Rp. 250.001 - Rp.500.000	0.09	0.05
	Rp. 500.001 - Rp.750.000	0.01	0.08
	Rp. 750.001 - Rp. 1.000.000	0.00	0.05
	'>Rp. 1.000.000	0.00	0.15
Luas Bangunan	'< 25 M2	0.57	0.15
	25-50 M2	0.21	0.23
	50-99 M2	0.06	0.30
	100-200 M2	0.07	0.13
	'>200 M2	0.09	0.20
MCK	Berbagi Pakai	0.07	0.08
	Kepemilikan Sendiri Diluar	0.37	0.10
	Kepemilikan Sendiri Didalam	0.55	0.83

Atribut		Terima	Tidak Terima
Pekerjaan Istri	Tidak Bekerja	0.78	0.65
	Petani	0.09	0.05
	Nelayan	0.00	0.00
	Peg. Swasta	0.00	0.00
	Wirausaha	0.00	0.08
	Lainnya	0.13	0.23

Adapun sebagian penjelasan terkait perhitungan dari Tabel 4.16 diatas sebagai berikut :

1. Menghitung menggunakan Metode *Naïve Bayes* Langkah awal adalah menghitung terlebih dahulu *Probabilitas* kelasnya. Sehingga perhitungan yang sesuai data di atas adalah sebagai berikut :

$$P(h_j|x) = \frac{p(x|h)p(h)}{p(x)}$$

$$P(\text{Status}|\text{Terima}) = \frac{67}{107} = 0.63$$

$$P(\text{Status}|\text{Tidak Terima}) = \frac{40}{107} = 0.37$$

2. Menghitung *Probabilitas* Atribut Pekerjaan KK pada Class Terima

$$P(\text{Pekerjaan KK} = \text{Tidak Bekerja}|\text{Terima}) = \frac{6}{67} = 0.9$$

$$P(\text{Pekerjaan KK} = \text{Petani}|\text{Terima}) = \frac{45}{67} = 0.67$$

$$P(\text{Pekerjaan KK} = \text{Nelayan}|\text{Terima}) = \frac{5}{67} = 0.07$$

$$P(\text{Pekerjaan KK} = \text{Peg. Swasta}|\text{Terima}) = \frac{1}{67} = 0.01$$

$$P(\text{Pekerjaan KK} = \text{Wirausaha}|\text{Terima}) = \frac{1}{67} = 0.01$$

$$P(\text{Pekerjaan KK} = \text{Lainnya}|\text{Terima}) = \frac{9}{67} = 0.13$$

3. Menghitung *Probabilitas* Atribut Pekerjaan KK pada Class Tidak Terima

$$P(\text{Pekerjaan KK} = \text{Tidak Bekerja}|\text{Tidak Terima}) = \frac{4}{40} = 0.10$$

$$P(\text{Pekerjaan KK} = \text{Petani}|\text{Tidak Terima}) = \frac{45}{40} = 0.08$$

$$P(\text{Pekerjaan KK} = \text{Nelayan}|\text{Tidak Terima}) = \frac{0}{40} = 0.0$$

$$P(\text{Pekerjaan KK} = \text{Peg. Swasta}|\text{Tidak Terima}) = \frac{2}{40} = 0.05$$



$$P(\text{Pekerjaan KK} = \text{Wirausaha} | \text{Terima}) = \frac{5}{40} = 0.13$$

$$P(\text{Pekerjaan KK} = \text{Lainnya} | \text{Tidak Terima}) = \frac{26}{40} = 0.65$$

Sehingga untuk melakukan perhitungan terhadap atribut lainnya sama seperti cara perhitungan di atas.

#### 4. Menguji Hasil Klasifikasi *Naïve Bayes Clasification*

Formula :  $P(c|F_1, F_2, F_3, \dots, F_n) = p(C) \prod_{i=1}^n p(F_i|C)$

Penyelesaian :

$$\begin{aligned} P(\text{Status} | \text{Terima}) &= P(\text{Status} | \text{Terima}) * (P(\text{Pekerjaan KK} | \text{Petani}) * \\ &(P(\text{Penghasilan KK} | \text{Rp. 250.001} - \text{Rp. 500.000}) * \\ &(P(\text{Penghasilan Istri} | \text{Tidak Berpenghasilan}) * (P(\text{Luas Tanah} | < \\ &25 \text{ M2}) * (P(\text{Luas Bangunan} | < 25 \text{ M2}) * \\ &(P(\text{Pekerjaan Istri} | \text{Tidak Bekerja}) * \\ &(P(\text{MCK} | \text{Kepemilikan Sendiri Didalam})) \\ &= 0.63 * (0.67 * 0.42 * 0.49 * 0.48 * 0.57 * 0.78 * 0.55) \\ &= 0.63 * 0.02 \\ &= 0.01 \\ P(\text{Status} | \text{Tidak Terima}) &= \\ P(\text{Status} | \text{Tidak Terima}) * (P(\text{Pekerjaan KK} | \text{Petani}) * \\ &(P(\text{Penghasilan KK} | \text{Rp. 250.001} - \text{Rp. 500.000}) * \\ &(P(\text{Penghasilan Istri} | \text{Tidak Berpenghasilan}) * (P(\text{Luas Tanah} | < \\ &25 \text{ M2}) * (P(\text{Luas Bangunan} | < 25 \text{ M2}) * \\ &(P(\text{Pekerjaan Istri} | \text{Tidak Bekerja}) * \\ &(P(\text{MCK} | \text{Kepemilikan Sendiri Didalam})) \\ &= 0.37 * (0.08 * 0.23 * 0.60 * 0.13 * 0.15 * 0.65 * 0.83) \\ &= 0.37 * 0.00 \\ &= 0.00 \end{aligned}$$

#### 4.2.4 Pengujian Model Dan Evaluasi Kinerja Model

Setelah dilakukan pemodelan Penerima Bantuan Langsung Tunai dengan metode *Naïve Bayes* maka perlu dilakukan pengujian model dengan menggunakan *accuracy score* dan evaluasi kinerja model menggunakan *Confusion Matrix* yang

terdiri dari *accuracy*, *recall*, *precision*, dan *f1-score* sebelum dan setelah dilakukan seleksi atribut untuk mengetahui Tingkat kinerja masing-masing model.

#### 4.2.5 Implementasi Hasil Pemodelan

Pada sub bab ini akan menjelaskan tentang tahapan implementasi dari uji coba Pemodelan untuk Klasifikasi Penerima Bantuan Langsung Tunai.

#### 4.2.6 Implementasi Data Processing

Sebelum dilakukan pemodelan ke dalam metode *Naïve Bayes* terlebih dahulu dilakukan beberapa pengolahan data sebagai berikut :

##### 1. *Pseudocode* Import Library Python

1.	<code>import pandas as pd</code>
2.	<code>import numpy as np</code>
3.	<code>from sklearn.feature_selection import SelectKBest</code>
4.	<code>from sklearn.feature_selection import mutual_info_classif</code>
5.	<code>from sklearn.naive_bayes import GaussianNB</code>
6.	<code>import matplotlib.pyplot as plt</code>
7.	<code>import seaborn as sns</code>
8.	<code>import plotly.graph_objs as go</code>

#### *Pseudocode 4. 1 Import Library Python*

Penjelasan mengenai library di atas ditunjukkan pada Tabel 4.18.

**Tabel 4. 18** Deskripsi Library

Library	Deskripsi
Pandas	Mengimpor library Pandas dengan alias pd Pandas adalah library untuk manipulasi dan analisis data, terutama dalam bentuk tabel (DataFrame).
Numpy	Mengimpor library NumPy dengan alias np. NumPy menyediakan dukungan untuk array dan operasi matematika tingkat tinggi,
Sklearn.feature_selection	Mengimpor kelas SelectKBest dari modul sklearn.feature_selection. SelectKBest digunakan untuk memilih fitur terbaik (dalam jumlah K) berdasarkan skor statistik tertentu dari dataset.
Mutual_info_classif	Mengimpor fungsi mutual_info_classif dari modul sklearn.feature_selection. Fungsi ini untuk menghitung informasi mutual antara fitur dan target



Library	Deskripsi
	untuk tugas klasifikasi. Informasi mutual adalah ukuran ketergantungan antara dua variabel.
GaussianNB	Mengimpor kelas GaussianNB dari modul sklearn.naive_bayes. GaussianNB adalah implementasi dari algoritma <i>Naive Bayes</i> dengan asumsi distribusi normal (Gaussian) dari fitur-fitur. Algoritma ini sering digunakan untuk klasifikasi.
matplotlib.pyplot	Mengimpor modul Pyplot dari library Matplotlib dengan alias plt. Matplotlib adalah library plotting yang sangat kuat di Python, dan Pyplot adalah antarmuka yang umum digunakan untuk membuat grafik dan plot.
seaborn	Mengimpor library Seaborn dengan alias sns. Seaborn adalah library visualisasi yang menyediakan antarmuka tingkat tinggi untuk membuat plot statistik yang menarik dan informatif.
plotly.graph_objs	Mengimpor modul graph_objs dari library Plotly dengan alias go. Plotly adalah library untuk membuat visualisasi interaktif yang sangat kaya dan dapat dipersonalisasi, dan graph_objs menyediakan objek grafik yang mendasari pembuatan grafik di Plotly.

## 2. Pseudocode menghubungkan Google Colab dengan Google Drive

1.	<code>from google.colab import drive drive.mount('/content/drive')</code>
----	---

### **Pseudocode 4.2** Menghubungkan *Google Colab* dengan *Google Drive*

Kode tersebut adalah untuk mengakses Google Drive dari Google Colab, yang merupakan platform cloud yang memungkinkan Anda menulis dan mengeksekusi kode Python di browser web.

## 3. Pseudocode Load Dataset

1.	<code>#Baca Dataset Bansos</code>
2.	<code>dfBansos = pd.read_excel('/content/drive/MyDrive/Skripsi Fahrun/Dataset Bantuan Sosial.xlsx')</code>
3.	<code>dfBansos.head()</code>

### **Pseudocode 4.3** Load Dataset

Kode tersebut bertujuan untuk membaca dataset yang tersimpan dalam format Excel dari Google Drive menggunakan Pandas, dan menampilkannya di dalam notebook Google Colab.

#### 4. Pseudocode Konversi Dataset Excel ke CSV

1.	#Konversi dataset Bansos Excel ke CSV
2.	dfBansos.to_csv("/content/drive/MyDrive/Skripsi Fahrun/DatasetBansos.csv", index=False)
3.	dfBansos.info()

#### ***Pseudocode 4. 4*** Konversi Dataset Excel ke CSV

Coding tersebut bertujuan untuk mengkonversi dataset yang telah dibaca sebelumnya dari format Excel menjadi format CSV, serta memberikan informasi tentang dataset tersebut.

#### 5. Pseudocode Menampilkan beberapa baris pertama

1.	dfBansos.head()
----	-----------------

#### ***Pseudocode 4. 5*** Menampilkan beberapa baris pertama

Coding tersebut bertujuan untuk menampilkan beberapa baris pertama dari dataset tersebut.

#### 6. Pseudocode mengecek jumlah kelas

1.	#Mengecek Jumlah untuk masing-masing kelas
2.	dfBansos['Status'].value_counts()

#### ***Pseudocode 4. 6*** Mengecek jumlah kelas

Coding tersebut bertujuan untuk menghitung jumlah kemunculan atau frekuensi dari setiap nilai unik di kolom Status pada DataFrame dfBansos.

#### 7. Pseudocode Mengecek jumlah Atribut Kosong

1.	#Mengecek Jumlah Atribut yg kosong
2.	dfBansos.isnull().sum()

#### ***Pseudocode 4. 7*** Mengecek jumlah Atribut Kosong

Coding tersebut bertujuan untuk menghitung serta menampilkan jumlah nilai kosong disetiap kolom dari DataFrame 'dfBansos'.

#### 8. Pseudocode mengecek jumlah atribut Pekerjaan KK

1.	<code>dfBansos['Pekerjaan KK'].value_counts()</code>
----	--

### **Pseudocode 4. 8** Mengecek jumlah atribut Pekerjaan KK

Coding tersebut bertujuan untuk menghitung jumlah kemunculan atau frekuensi dari setiap nilai unik di kolom Pekerjaan KK pada DataFrame dfBansos.

#### 9. Pseudocode Konversi Data String ke Angka

1.	<code># Konversi Data String ke Angka</code>
2.	<code>dfBansos['Pekerjaan KK'].replace(['Tidak</code>
3.	<code>Bekerja','Petani','Nelayan','Peg. Swasta',</code>
	<code>'Wirausaha','Lainnya'],</code>
4.	<code>[1,2,3,4,5,6], inplace=True)</code>
5.	<code>dfBansos.head()</code>

### **Pseudocode 4. 9** Konversi Data String ke Angka

Coding tersebut Untuk mengkonversi nilai-nilai kategori dalam kolom Pekerjaan KK pada DataFrame dfBansos menjadi nilai numerik.

#### 10. Pseudocode Simpan Hasil Konversi

1.	<code>#Simpan Hasil Konversi</code>
2.	<code>dfBansos.to_excel("/content/drive/MyDrive/Skripsi</code>
	<code>Fahrun/DatasetBansosAngka.xlsx", index=False)</code>

### **Pseudocode 4. 10** Simpan Hasil Konversi

Fungsi dari kode tersebut adalah untuk menyimpan DataFrame dfBansos ke dalam sebuah file Excel

#### 11. Pseudocode Konversi Dataset Bansos Excel ke CSV

1.	<code>#Konversi dataset Bansos Excel ke CSV</code>
2.	<code>dfBansos.to_csv("/content/drive/MyDrive/Skripsi</code>
	<code>Fahrun/DatasetBansosAngka.csv", index=False)</code>
3.	<code>dfBansos.info()</code>

### **Pseudocode 4. 11** Konversi Dataset Bansos Excel ke CSV

Kode tersebut melakukan dua fungsi utama: mengonversi dataset dari format Excel ke format CSV dan kemudian menampilkan informasi tentang dataset.

#### 12. *Pseudocode* Konversi Dataset Bansos Excel ke CSV

```
1. #Load dataset CSV setelah run 1 dan 2 langsung dari sini
2. dfBansos=pd.read_csv("/content/drive/MyDrive/Skripsi
   Fahrun/DatasetBansosAngka.csv")
3. dfBansos.info()
```

#### **Pseudocode 4. 12** Konversi Dataset Bansos Excel ke CSV

Coding tersebut bertujuan untuk membaca data dari file CSV dan menampilkan informasi tentang struktur data tersebut.

#### 13. *Pseudocode* Setting Dataset Atribut dengan nilai skor yang tinggi

```
1. #setting dataset dengan nilai skor yang tinggi
2. X = dfBansos[['Pekerjaan KK','Pekerjaan Istri',
   'Penghasilan KK','Penghasilan Istri',
   'Jumlah Tanggungan','Kepemilikan Rumah',
   'Sumber Listrik','Luas Tanah','Luas Bangunan',
   'Sumber Air','MCK']].values #variabel input
3. y = dfBansos[['Status']].values #varibael target/output
```

#### **Pseudocode 4. 13** Setting dataset atribut dengan nilai skor yang tinggi

Coding tersebut bertujuan untuk menyiapkan dataset untuk analisis atau pembelajaran mesin.

#### 14. *Pseudocode* membagi data Training dan Testing

```
1. # Train test split untuk membagi data training dan testing
2. from sklearn.model_selection import train_test_split
3. X_train, X_test, y_train, y_test = train_test_split(X, y,
   test_size=0.2, random_state=10)
4. print ('Train set:', X_train.shape, y_train.shape)
5. print ('Test set:', X_test.shape, y_test.shape)
```

#### **Pseudocode 4. 14** Membagi data Training dan Testing



Coding tersebut berfungsi untuk membagi dataset menjadi set pelatihan (*training set*) dan set pengujian (*testing set*).

#### 15. Pseudocode Naïve Bayes untuk semua atribut

```

1. from sklearn.metrics import accuracy_score
2. # Train Naive Bayes classifier
3. nb_classifier = GaussianNB()
4. nb_classifier.fit(X_train, y_train)

5. # Predict on test set
6. y_pred = nb_classifier.predict(X_test)

7. # Calculate accuracy
8. accuracy = accuracy_score(y_test, y_pred)
9. print("Accuracy:", accuracy)

```

#### **Pseudocode 4. 15** Naïve Bayes untuk semua atribut

Coding tersebut berfungsi untuk melatih model *Naive Bayes* serta menghitung akurasi dengan menggunakan dataset yang sudah dibagi sebelumnya menjadi data training serta data testing.

#### 16. Pseudocode Confusion Matrix

```

1. from sklearn.metrics import confusion_matrix,
2. classification_report
3. import seaborn as sns
4. import matplotlib.pyplot as plt

5. # Evaluasi Model
6. y_pred = nb_classifier.predict(X_test)

7. # Membuat confusion matrix
8. cm = confusion_matrix(y_test, y_pred)

9. # Mencetak confusion matrix
10. labels = ['Terima', 'Tidak Terima']
11. plt.figure(figsize=(5, 4))
12. sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels,
13. yticklabels=labels, cmap='viridis')
14. plt.xlabel('Klasifikasi')
15. plt.ylabel('Actual')
16. plt.title('Confusion Matrix')

```

17.	<code>plt.show()</code>
18.	
19.	<code># Menghitung precision, recall, f1-score, dan support</code>
20.	<code>report = classification_report(y_test, y_pred,</code>
21.	<code>target_names=labels)</code>
22.	<code>print(report)</code>

#### Pseudocode 4. 16 Confusion Matrix

Coding tersebut bertujuan untuk mengevaluasi model Klasifikasi, seperti model *Naive Bayes*, dengan menggunakan metrimm evaluasi, yaitu *Confusion Matrix*, *precision*, *recall*, dan *f1 Score*.

17. *Pseudocode* Setting dataset 7 atribut dengan nilai skor yang tinggi

1.	<code>#setting dataset dengan nilai skor yang tinggi 7 Atribut</code>
2.	<code>X = dfBansos[['Penghasilan KK', 'Penghasilan</code>
3.	<code>Istri', 'Pekerjaan KK',</code>
	<code>          'Luas Bangunan', 'Luas Tanah', 'Sumber Air',</code>
	<code>          'MCK']].values #variabel input</code>
4.	<code>y = dfBansos[['Status']].values #varibael target/output</code>

#### Pseudocode 4. 17 Setting dataset 7 atribut dengan nilai skor yang tinggi

Coding tersebut bertujuan untuk mempersiapkan dataset dengan memisahkan atribut yang relevan sebagai variabel input (x) dan kolom target sebagai variabel output (y). Langkah ini adalah bagian dari preprocessing data, yang penting sebelum data dimasukkan ke dalam algoritma pembelajaran mesin untuk pelatihan atau prediksi.

18. *Pseudocode* Membagi data Training dan Testing

1.	<code># Train test split untuk membagi data training dan testing</code>
2.	<code>from sklearn.model_selection import train_test_split</code>
3.	<code>X_train, X_test, y_train, y_test = train_test_split(X, y,</code>
4.	<code>test_size=0.2, random_state=10)</code>
5.	<code>print ('Train set:', X_train.shape, y_train.shape)</code>
6.	<code>print ('Test set:', X_test.shape, y_test.shape)</code>

#### Pseudocode 4. 18 Membagi data Training dan Testing

Kode ini berfungsi untuk membagi dataset menjadi dua bagian: satu bagian untuk melatih model ( $X_{\text{train}}$ ,  $y_{\text{train}}$ ) dan satu bagian untuk menguji model ( $X_{\text{test}}$ ,  $y_{\text{test}}$ ). Pembagian ini penting untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya (data pengujian) dan untuk memastikan model tidak overfitting pada data pelatihan.

#### 19. Pseudocode Naïve Bayes untuk 7 atribut

```

1. from sklearn.metrics import accuracy_score
2. # Train Naïve Bayes classifier
3. nb_classifier = GaussianNB()
4. nb_classifier.fit(X_train, y_train)

5. # Predict on test set
6. y_pred = nb_classifier.predict(X_test)

7. # Calculate accuracy
8. accuracy = accuracy_score(y_test, y_pred)
9. print("Accuracy:", accuracy)

```

#### **Pseudocode 4. 19 Naïve Bayes untuk 7 atribut**

Coding tersebut berfungsi untuk melatih model *Naïve Bayes* pada 7 Atribut serta menghitung akurasi dengan menggunakan dataset yang sudah dibagi sebelumnya menjadi data training serta data testing.

#### 20. Pseudocode Confusion Matrix untuk 7 atribut

```

1. from sklearn.metrics import confusion_matrix,
2. classification_report
3. import seaborn as sns
4. import matplotlib.pyplot as plt

5. # Evaluasi Model
6. y_pred = nb_classifier.predict(X_test)

7. # Membuat confusion matrix
8. cm = confusion_matrix(y_test, y_pred)

9. # Mencetak confusion matrix
10. labels = ['Terima', 'Tidak Terima']
11. plt.figure(figsize=(5, 4))
12.

```



```

13. sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels,
14. yticklabels=labels, cmap='viridis')
15. plt.xlabel('Klasifikasi')
16. plt.ylabel('Actual')
17. plt.title('Confusion Matrix')
18. plt.show()

19. # Menghitung precision, recall, f1-score, dan support
20. report = classification_report(y_test, y_pred,
21. target_names=labels)
22. print(report)

```

#### **Pseudocode 4. 20** *Confusion Matrix* untuk 7 atribut

Coding tersebut bertujuan untuk mengevaluasi model Klasifikasi, seperti model *Naive Bayes*, dengan menggunakan metrimm evaluasi, yaitu *Confusion Matrix*, *precision*, *recall*, dan *f1 Score*.

#### 21. *Pseudocode* Setting dataset 9 atribut dengan nilai skor yang tinggi

```

1. #setting dataset dengan nilai skor yang tinggi 9 Atribut
2. X = dfBansos[['Penghasilan KK', 'Penghasilan
3. Istri', 'Pekerjaan KK',
           'Luas Bangunan', 'Luas Tanah', 'MCK', 'Sumber
4. Air', 'Sumber Listrik',
           'Pekerjaan Istri']].values #variabel input
y = dfBansos[['Status']].values #varibael target/output

```

#### **Pseudocode 4. 21** Setting dataset 9 atribut dengan nilai skor yang tinggi

Coding tersebut bertujuan untuk mempersiapkan dataset dengan memisahkan atribut yang relevan sebagai variabel input (x) dan kolom target sebagai variabel output (y). Langkah ini adalah bagian dari *Preprocessing* data, yang penting sebelum data dimasukkan ke dalam algoritma pembelajaran mesin untuk pelatihan atau prediksi.

#### 22. *Pseudocode* Membagi data Training dan Testing

```

1. # Train test split untuk membagi data training dan testing
2. from sklearn.model_selection import train_test_split

```

```

3. X_train, X_test, y_train, y_test = train_test_split(X, y,
4. test_size=0.2, random_state=10)

5. print ('Train set:', X_train.shape, y_train.shape)
6. print ('Test set:', X_test.shape, y_test.shape)

```

#### **Pseudocode 4. 22** Membagi data Training dan Testing

Kode ini berfungsi untuk membagi dataset menjadi dua bagian: satu bagian untuk melatih model (X\_train, y\_train) dan satu bagian untuk menguji model (X\_test, y\_test). Pembagian ini penting untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya (data pengujian) dan untuk memastikan model tidak overfitting pada data pelatihan.

#### **23. Pseudocode Naïve Bayes** untuk 9 atribut

```

1. from sklearn.metrics import accuracy_score
2. # Train Naive Bayes classifier
3. nb_classifier = GaussianNB()
4. nb_classifier.fit(X_train, y_train)

5. # Predict on test set
6. y_pred = nb_classifier.predict(X_test)

7. # Calculate accuracy
8. accuracy = accuracy_score(y_test, y_pred)
9. print("Accuracy:", accuracy)

```

#### **Pseudocode 4. 23** Naïve Bayes untuk 9 atribut

Coding tersebut berfungsi untuk melatih model *Naive Bayes* pada 9 Atribut serta menghitung akurasi dengan menggunakan dataset yang sudah dibagi sebelumnya menjadi data training serta data testing.

#### **24. Pseudocode Confusion Matrix** untuk 9 atribut

```

1. from sklearn.metrics import confusion_matrix,
2. classification_report
3. import seaborn as sns
4. import matplotlib.pyplot as plt

5. # Evaluasi Model
6. y_pred = nb_classifier.predict(X_test)

```

```

7. # Membuat confusion matrix
8. cm = confusion_matrix(y_test, y_pred)

9. # Mencetak confusion matrix
10. labels = ['Terima', 'Tidak Terima']
11. plt.figure(figsize=(5, 4))
12. sns.heatmap(cm, annot=True, fmt='d', xticklabels=labels,
13. yticklabels=labels, cmap='viridis')
14. plt.xlabel('Klasifikasi')
15. plt.ylabel('Actual')
16. plt.title('Confusion Matrix')
17. plt.show()

18.
19. # Menghitung precision, recall, f1-score, dan support
20. report = classification_report(y_test, y_pred,
21. target_names=labels)
22. print(report)

```

**Pseudocode 4. 24** *Confusion Matrix* untuk 9 atribut

Coding tersebut bertujuan untuk mengevaluasi model Klasifikasi, seperti model *Naive Bayes*, dengan menggunakan metrim evaluasi, yaitu *Confusion Matrix*, *precision*, *recall*, dan *f1 Score*.

## BAB V PEMBAHASAN PENELITIAN

### 5.1 Pembahasan Kinerja Model

Pemodelan *Naive Bayes* dalam klasifikasi Penerima Bantuan Langsung Tunai dilakukan uji coba pemodelan sebanyak 2 kali, yaitu sebelum diseleksi atribut dan setelah seleksi atribut menggunakan metode *Information Gain*, dimana kinerja model tersebut diukur dengan menggunakan *Confusion Matrix* yang selanjutnya di ukur masing-masing item seperti : *Precision*, *Recall*, *F1-Score*, dan *Accuracy*. Hasil masing-masing percobaan direkap dalam bentuk tabel yang ditunjukkan pada Tabel 5.1 berikut :

**Tabel 5. 1** Kinerja Model *Naive Bayes*

Uji Coba	Status	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Jumlah Data
Sebelum Seleksi Atribut ( 11 Atribut)	Terima	0.84	0.94	0.70	17
	Tidak Terima	0.88	0.70	0.78	10
	Akurasi	<b>0.85</b>			
Setelah Seleksi Atribut ( 3 Atribut)	Terima	0.77	1.00	0.87	17
	Tidak Terima	1.00	0.50	0.67	10
	Akurasi	<b>0.81</b>			
Setelah Seleksi Atribut ( 7 Atribut)	Terima	0.94	1.00	0.97	17
	Tidak Terima	1.00	0.90	0.95	10
	Akurasi	<b>0.96</b>			
Setelah Seleksi Atribut ( 9 Atribut)	Terima	0.89	1.00	0.94	17
	Tidak Terima	1.00	0.80	0.89	10
	Akurasi	<b>0.93</b>			

Berdasarkan Tabel 5.1 di atas ditunjukkan bahwa akurasi *Naive Bayes* dalam klasifikasi Penerima Bantuan Langsung Tunai sebelum dilakukan seleksi atribut (11 atribut) sebesar 85%, dan setelah dilakukan seleksi atribut penting dengan *Information Gain* pada 3 atribut didapatkan hasil akurasi sebesar 81%, sedangkan untuk 7 atribut sebesar 96% dan pada 9 atribut sebesar 93 %. Dari hasil ini dapat disimpulkan bahwa untuk mendapatkan tingkat akurasi yang bagus adalah dengan menggunakan 7 atribut.

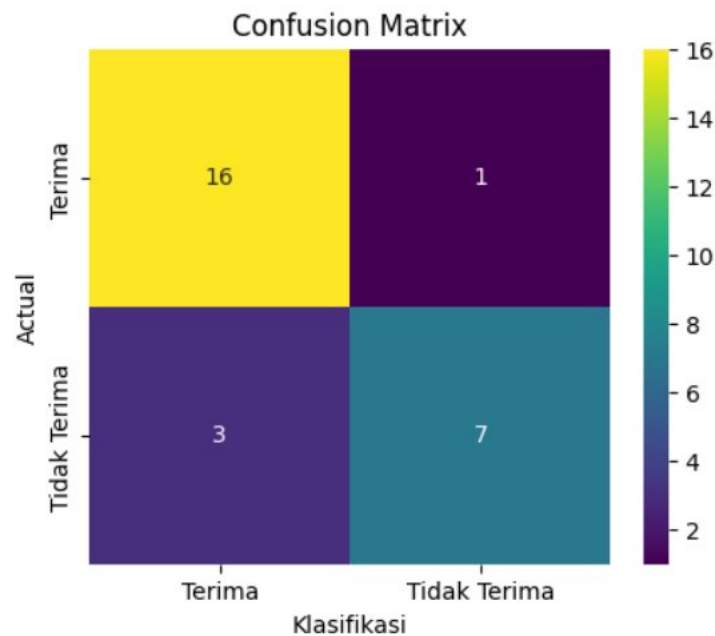
## **5.2 Pembahasan Hasil Pemodelan**

Pencapaian kedua tujuan penelitian yaitu Menganalisis hasil seleksi atribut penting pada Penerapan *Information Gain* Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan *Naïve Bayes* dan Menerapkan *Infomation Gain* dalam seleksi atribut penting pada *Naive Bayes*, sehingga didapatkan hasil akurasi yang lebih akurat, maka perlu diuraikan pembahasan hasil pemodelan.

### **5.2.1 Hasil Pemodelan *Naïve Bayes* Sebelum Seleksi Atribut**

Pemodelan *Naive Bayes* sebelum seleksi atribut perlu dilakukan supaya hasilnya dapat dibandingkan mana hasil klasifikasi yang lebih akurat, apakah sebelum seleksi atribut atau setelah seleksi atribut. Untuk mengetahui atau mengukur akurasi hasil suatu klasifikasi menggunakan teknik seperti *Precision*, *Recall*, *F1-Score*, dan *Accuracy* yang didapatkan dari hasil perhitungan *Confusion Matrix*. Hasil perhitungan *Confusion Matrix* ditunjukkan pada Gambar 5.1 berikut :





**Gambar 5. 1** Hasil Perhitungan *Confusion Matrix*

Selanjutnya dilakukan perhitungan *Precision*, *Recall*, *F1-Score*, dan *Accuracy* berdasarkan Gambar 5.1 Hasil perhitungannya ditunjukkan pada Tabel 5.2.

**Tabel 5. 2** *Classification Report* Klasifikasi Sebelum Seleksi Atribut

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Terima	0.84	0.94	0.70
Tidak Terima	0.88	0.70	0.78
Akurasi	<b>0.85</b>		

Berdasarkan Tabel 5.2 di atas, ditunjukkan bahwa didapatkan nilai akurasi sebesar 85 %. *Precision*, *Recall*, *F1-Score*, dan *Accuracy* dapat dihitung menggunakan persamaan berikut :

1. *Precision*

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{16}{16+3} = \frac{16}{19} = 0.84$$

2. *Recall*



$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{16}{16+1} = \frac{16}{17} = 0.94$$

### 3. *F1-Score*

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

$$F1 = 2 \times \frac{0.84 \times 0.94}{0.84+0.94} = 2 \times \frac{0.79}{1.78} = 0.89$$

### 4. *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{16+7}{16+7+3+1} = \frac{23}{27} = 0.85$$

Berdasarkan perhitungan diatas maka berikut dapat diuraikan penjelasan terkait perhitungan tersebut :

#### 1. *Precision* = 0.84

*Precision* adalah proporsi prediksi positif yang benar-benar positif. Artinya, dari semua prediksi yang diklasifikasikan sebagai positif oleh model, 84% di antaranya benar-benar positif, Ini menunjukkan seberapa baik model menghindari kesalahan dalam mengklasifikasikan negatif sebagai positif.

#### 2. *Recall* = 0,94

*Recall* adalah proporsi kasus positif yang benar yang diidentifikasi dengan benar oleh model. Artinya, dari semua kasus yang sebenarnya positif, model dapat mengidentifikasi 94% di antaranya dengan benar, ini menunjukkan seberapa baik model menemukan semua kasus positif dalam dataset.

#### 3. *F1-Score* = 0.89

*F1-Score* adalah rata-rata harmonis dari *Precision* dan *Recall*. Nilai ini memberikan gambaran umum tentang keseimbangan antara *Precision* dan recall. Dengan nilai F1-Score sebesar 0.89, ini menunjukkan bahwa model memiliki keseimbangan yang baik antara mengidentifikasi kasus positif dengan benar (recall) dan menghindari kesalahan klasifikasi negatif sebagai positif (precision).

#### 4. *Accuracy* = 0.85

*Accuracy* adalah proporsi dari semua prediksi yang benar (baik positif maupun negatif) terhadap total jumlah sampel. Artinya, model mengklasifikasikan 85% dari seluruh sampel dengan benar. Ini memberikan gambaran umum tentang kinerja model, tetapi mungkin tidak mencerminkan kinerja yang sebenarnya jika dataset tidak seimbang (misalnya, jika satu kelas jauh lebih dominan daripada kelas lainnya).

#### 5.2.2 Hasil Seleksi Atribut Penting dengan *Information Gain*

Setelah dilakukan seleksi atribut penting pada dataset 134 sebanyak 11 atribut dengan menggunakan metode *Information Gain* maka didapatkan hasil skor kepentingan untuk masing-masing atribut yang ditunjukkan pada Gambar 5.2 berikut :

	Atribut	Score
2	Penghasilan KK	0.200365
3	Penghasilan Istri	0.197497
0	Pekerjaan KK	0.167354
7	Luas Tanah	0.124624
8	Luas Bangunan	0.083985
10	MCK	0.061642
1	Pekerjaan Istri	0.013110
4	Jumlah Tanggungan	0.000000
5	Kepemilikan Rumah	0.000000
6	Sumber Listrik	0.000000
9	Sumber Air	0.000000

**Gambar 5. 2** Hasil Skor Kepentingan

Berdasarkan Gambar 5.2 di atas hasil proses melalui tools Python, agar memudahkan dalam analisis dalam menentukan skor kepentingan masing-masing atribut diuraikan dalam bentuk Tabel seperti yang ditunjukkan pada Tabel 5.3 berikut :

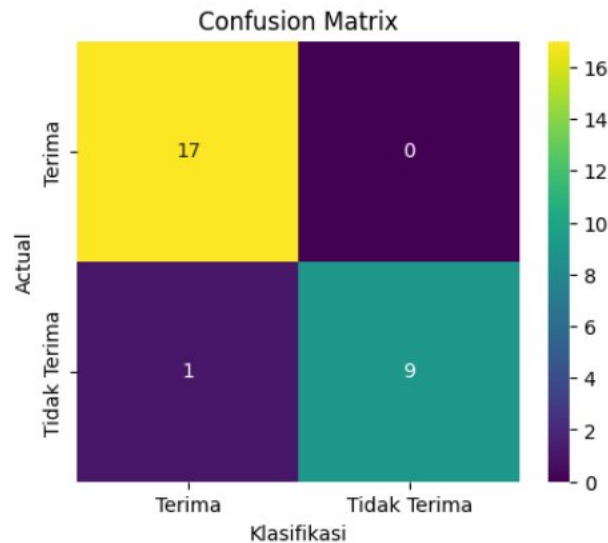
**Tabel 5. 3** Skor Kepentingan

NO	ATRIBUT	SCORE
1	Penghasilan KK	0.200365
2	Pekerjaan KK	0.197497
3	Penghasilan Istri	0.167354
4	Luas Tanah	0.124624
5	Luas Bangunan	0.083985
6	MCK	0.061642
7	Pekerjaan Istri	0.013110
8	Jumlah Tanggungan	0.000000
9	Kepemilikan Rumah	0.000000
10	Sumber Listrik	0.000000
11	Sumber Air	0.000000

Berdasarkan Tabel 5.3 di atas, dapat dilihat bahwa terdapat 11 atribut yang memiliki pengaruh terhadap hasil klasifikasi, namun hanya dipilih atribut yang memiliki skor kepentingan sebesar 0,01 ke atas yaitu atribut Penghasilan KK dengan skor 0.200365, Pekerjaan KK dengan skor 0.197497, Penghasilan Istri dengan skor 0.167354, Luas Tanah dengan skor 0.124624, Luas Bangunan dengan skor 0.083985, MCK dengan skor 0.061642 dan Pekerjaan Istri dengan skor 0.013110. atribut inilah yang akan dijadikan klasifikasi dengan metode *Naive Bayes*.

### **5.2.3 Hasil Pemodelan *Naive Bayes* Setelah Seleksi Atribut**

Setelah dilakukan seleksi atribut yang memiliki skor kepentingan tertinggi, dipilih sebanyak 7 atribut terpenting yaitu atribut Penghasilan KK, Pekerjaan KK, Penghasilan Istri, Luas Tanah, Luas Bangunan, MCK dan Pekerjaan Istri untuk menghitung kembali akurasi hasil klasifikasi digunakan cara yang sama seperti sebelumnya. Hasil perhitungan *Confusion Matrix*.ditunjukkan pada Gambar 5.3 berikut :



**Gambar 5. 3** Hasil Perhitungan *Confusion Matrix*

Selanjutnya dilakukan perhitungan *Precision*, *Recall*, *F1-Score*, dan Akurasi berdasarkan Gambar 5.3 Hasil perhitungannya ditunjukkan pada Tabel 5.4.

**Tabel 5. 4** Classification Report Klasifikasi Setelah Seleksi Atribut

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Terima	0.94	1.00	0.97
Tidak Terima	1.00	0.95	0.95
Akurasi	<b>0.96</b>		

Berdasarkan Tabel 5.2 di atas, ditunjukkan bahwa didapatkan nilai akurasi sebesar 96 %. *Precision*, *Recall*, *F1-Score*, dan *Accuracy* dapat dihitung menggunakan persamaan berikut :

1. *Precision*

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{17}{17+1} = \frac{17}{18} = 0.94$$

2. *Recall*

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{17}{17+0} = \frac{17}{17} = 1.00$$

3. *F1-Score*

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 = 2 \times \frac{0.94 \times 1.00}{0.94 + 1.00} = 2 \times \frac{0.94}{1.94} = 0.97$$

#### 4. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{17 + 9}{17 + 9 + 1 + 0} = \frac{26}{27} = 0.96$$

Berdasarkan perhitungan diatas maka berikut dapat diuraikan penjelasan terkait perhitungan tersebut :

##### 1. Precision = 0.94

*Precision* adalah proporsi prediksi positif yang benar-benar positif. Artinya, dari semua prediksi yang diklasifikasikan sebagai positif oleh model, 94% di antaranya benar-benar positif, Ini menunjukkan seberapa baik model menghindari kesalahan dalam mengklasifikasikan negatif sebagai positif.

##### 2. Recall = 1.00

*Recall* adalah proporsi kasus positif yang benar yang diidentifikasi dengan benar oleh model. Artinya, dari semua kasus yang sebenarnya positif, model dapat mengidentifikasi 100% di antaranya dengan benar, ini menunjukkan seberapa baik model menemukan semua kasus positif dalam dataset.

##### 3. F1-Score = 0.97

*F1-Score* adalah rata-rata harmonis dari *Precision* dan *Recall*. Nilai ini memberikan gambaran umum tentang keseimbangan antara *Precision* dan *Recall*. Dengan nilai *F1-Score* sebesar 0.97, ini menunjukkan bahwa model memiliki keseimbangan yang baik antara mengidentifikasi kasus positif dengan benar (*Recall*) dan menghindari kesalahan klasifikasi negatif sebagai positif (*Precision*).

##### 4. Accuracy = 0.96

*Accuracy* adalah proporsi dari semua prediksi yang benar (baik positif maupun negatif) terhadap total jumlah sampel. Artinya, model mengklasifikasikan 96% dari seluruh sampel dengan benar. Ini memberikan

gambaran umum tentang kinerja model, tetapi mungkin tidak mencerminkan kinerja yang sebenarnya jika dataset tidak seimbang (misalnya, jika satu kelas jauh lebih dominan daripada kelas lainnya).

Berdasarkan Tabel 5.2 dan Taabel 5.4 di atas, ditunjukkan bahwa sebelum seleksi atribut didapatkan akurasi sebesar 85 % dan setelah seleksi atribut didapatkan akurasi sebesar 96 %. hal ini menunjukkan bahwa terjadi peningkatan akurasi sebesar 11 % dengan demikian hasil klasifikasi setelah dilakukan seleksi atribut dengan *Information Gain* didapatkan hasil klasifikasi lebih baik atau lebih akurat.



## **BAB VI**

### **KESIMPULAN DAN SARAN**

#### **6.1 Kesimpulan**

1. Hasil analisis seleksi atribut penting pada Penerapan *Information Gain* Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan Naïve Bayes dari 11 atribut yakni Pekerjaan KK, Pekerjaan Istri, Penghasilan KK, Penghasilan Istri, Jumlah Tanggungan, Kepemilikan Rumah, Sumber Listrik, Luas Tanah, Luas Bangunan, Sumber Air, MCK didapatkan 7 atribut penting yaitu Penghasilan KK, Penghasilan Istri, Pekerjaan KK, Luas Tanah, Luas Bangunan, MCK dan Pekerjaan Istri.
2. Hasil akurasi sebelum seleksi atribut dengan *Information Gain* didapatkan nilai akurasi sebesar 85 % setelah seleksi atribut penting didapatkan nilai akurasi sebesar 96 %, didapatkan hasil bahwa nilai akurasi setelah penerapan *Information Gain* lebih besar sehingga ini menunjukkan bahwa hasil klasifikasi yang dihasilkan lebih akurat sebesar 11 %.

#### **6.2 Saran**

Berdasarkan kesimpulan di atas, saran kepada peneliti selanjutnya, yaitu:

1. Penelitian berikutnya dapat menggunakan metode seleksi atribut yang lainnya agar bisa dibandingkan hasil seleksi atribut terpenting
2. Dapat menggunakan metode klasifikasi yang lain untuk bisa dibandingkan hasilnya mana yang lebih baik.

## DAFTAR PUSTAKA

- [1] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, “Metode-metode Klasifikasi,” *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, p. 134, 2018.
- [2] D. Utami and P. A. R. Devi, “Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Metode Weighted Naïve Bayes Dengan Laplace Smoothing,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1373–1384, 2022, doi: 10.29100/jupi.v7i4.3592.
- [3] H. Nurrifqi *et al.*, “Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes,” vol. 10, no. 1, pp. 683–695, 2023.
- [4] A. Surahman, U. Hayati, B. Sosial, and R. Miner, “Implementasi Algoritma Naïve Bayes,” vol. 7, no. 1, pp. 347–352, 2023.
- [5] F. R. Sari, “Perbandingan Metode Naive Bayes Dan Regresi Logistik Biner Pada Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan(PKH),” 2023.
- [6] B. Pratama and A. S. Akbar, “Klasifikasi Penentuan Warga Penerima Bantuan Sosial di Masa Pandemi Menggunakan Metode Naive Bayes (Studi Kasus: Rt 002 Rw 01 Kel.Jagakarsa Kec.Jagakarsa Jakarta Selatan),” *J. Nas. Inform.*, vol. 2, no. 2, pp. 143–157, 2021, [Online]. Available:<http://ejournalibik57.ac.id/index.php/junif/article/view/338%0Ahttps://ejournal-ibik57.ac.id/index.php/junif/article/download/338/240>
- [7] D. A. Muthia, “Sentiment Analysis on Closure of Illegal Movie Streaming Sites Using Naïve Bayes Algorithm,” *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 123–128, 2020, doi: 10.33480/pilar.v16i1.1306.
- [8] A. Bengnga and R. Ishak, “Implementasi Seleksi Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Correlation Matrix with Heatmap,” *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 2, pp. 169–174, 2022, doi: 10.37905/jjee.v4i2.14403.

- [9] M. Norhalimi and T. A. Y. Siswa, "Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 3, pp. 237–255, 2022, doi: 10.14421/jiska.2022.7.3.237-255.
- [10] B. N. Sari, "Implementasi teknik seleksi fitur information gain pada algoritma klasifikasi machine learning untuk prediksi performa akademik siswa," *Semin. Nas. Teknol. Inf. dan Multimed. 2016*, pp. 55–60, 2016.
- [11] Humas Kominfo, "Program Bantuan Sosial Untuk Rakyat," *Kominfo*, 2018. [https://www.kominfo.go.id/index.php/content/detail/15708/program-bantuan-sosial-untuk-rakyat/0/artikel\\_gpr](https://www.kominfo.go.id/index.php/content/detail/15708/program-bantuan-sosial-untuk-rakyat/0/artikel_gpr) (accessed Sep. 10, 2023).
- [12] Muhammad Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. lhokseumawe: Andi Offset, 2020.
- [13] D. A. Pratiwi, R. M. Awangga, and M. Y. H. Setyawan, *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes*. Bandung: Kreatif Industri Nusantara, 2020.
- [14] Endang Etriyanti, "Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa," *J. Ilm. Bin. STMIK Bina Nusantara. Jaya Lubuklinggau*, vol. 3, no. 1, pp. 6–14, 2021, doi: 10.52303/jb.v3i1.40.
- [15] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan," *J. TIM Darmajaya*, vol. 02, no. 01, pp. 52–62, 2016.

## LAMPIRAN 1 KODE PROGRAM

```

import pandas as pd
import numpy as np
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import mutual_info_classif
from sklearn.naive_bayes import GaussianNB

import matplotlib.pyplot as plt

import seaborn as sns
import plotly.graph_objs as go

```

```

[ ] from google.colab import drive
    drive.mount('/content/drive')

```

Mounted at /content/drive

```

[ ] #Baca Dataset Bansos
    dfBansos = pd.read_excel('/content/drive/MyDrive/Skripsi Fahrur/Dataset Bantuan Sosial.xlsx')
    dfBansos.head()

```

	No	Nama KK	Pekerjaan KK	Pekerjaan Istri	Penghasilan KK	Penghasilan Istri
0	1	Alimin Dunggio	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan
1	2	M. TAKDIR	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan
2	3	Erwin S.M.Mosey	Petani	Tidak Bekerja	Rp. 500.001 - Rp. 750.000	Tidak Berpenghasilan
3	4	ALPIAN POTABUGA	Petani	Tidak Bekerja	Rp. 250.001 - Rp. 500.000	Tidak Berpenghasilan
4	5	Zainal Abidin	Petani	Lainnya	Rp. 250.001 - Rp. 500.000	>Rp. 1.000.000

```

[ ] #Konversi dataset Bansos Excel ke CSV
    dfBansos.to_csv("/content/drive/MyDrive/Skripsi Fahrur/DatasetBansos.csv", index=False)
    dfBansos.info()

```

```
#Konversi dataset Bansos Excel ke CSV
dfBansos.to_csv("/content/drive/MyDrive/Skripsi Fahrur/DatasetBansos.csv", index=False)
dfBansos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134 entries, 0 to 133
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   No                     134 non-null   int64
1   Nama KK                134 non-null   object
2   Pekerjaan KK           134 non-null   object
3   Pekerjaan Istri        134 non-null   object
4   Penghasilan KK         134 non-null   object
5   Penghasilan Istri     134 non-null   object
6   Jumlah Tanggungan     134 non-null   object
7   Kepemilikan Rumah     134 non-null   object
8   Sumber Listrik        134 non-null   object
9   Luas Tanah            134 non-null   object
10  Luas Bangunan          134 non-null   object
11  Sumber Air            134 non-null   object
12  MCK                   134 non-null   object
13  Status                134 non-null   object
dtypes: int64(1), object(13)
memory usage: 14.8+ KB
```

```
[ ] dfBansos.head()
```

```

      No      Nama KK  Pekerjaan KK  Pekerjaan Istri  Penghasilan KK  Penghasilan Istri
0  1  Alimin Dunggio      Petani  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
1  2      M. TAKDIR      Petani  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
2  3  Erwin S.M.Mosey      Petani  Tidak Bekerja  Rp. 500.001 - Rp. 750.000  Tidak Berpenghasilan
3  4  ALPIAN POTABUGA      Petani  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
4  5      Zainal Abidin      Petani      Lainnya  Rp. 250.001 - Rp. 500.000      >Rp. 1.000.000
```

```
[ ] # Konversi Data String ke Angka
dfBansos['Pekerjaan KK'].replace(['Tidak Bekerja','Petani','Nelayan','Peg. Swasta', 'Wirausaha','La
                                [1,2,3,4,5,6], inplace=True)
dfBansos.head()
```

```

      No      Nama KK  Pekerjaan KK  Pekerjaan Istri  Penghasilan KK  Penghasilan Istri
0  1  Alimin Dunggio      2  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
1  2      M. TAKDIR      2  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
2  3  Erwin S.M.Mosey      2  Tidak Bekerja  Rp. 500.001 - Rp. 750.000  Tidak Berpenghasilan
3  4  ALPIAN POTABUGA      2  Tidak Bekerja  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
4  5      Zainal Abidin      2      Lainnya  Rp. 250.001 - Rp. 500.000      >Rp. 1.000.000
```

```
dfBansos['Pekerjaan Istri'].value_counts()
```

```

Pekerjaan Istri
Tidak Bekerja    92
Lainnya          30
Petani           9
Wirausaha        3
Name: count, dtype: int64
```

```
[ ] # Konversi Data String ke Angka
dfBansos['Pekerjaan Istri'].replace(['Tidak Bekerja','Petani', 'Wirausaha','Lainnya'],
                                   [1,2,3,4], inplace=True)
dfBansos.head()
```

```

      No      Nama KK  Pekerjaan KK  Pekerjaan Istri  Penghasilan KK  Penghasilan Istri
0  1  Alimin Dunggio      2      1  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
1  2      M. TAKDIR      2      1  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
2  3  Erwin S.M.Mosey      2      1  Rp. 500.001 - Rp. 750.000  Tidak Berpenghasilan
3  4  ALPIAN POTABUGA      2      1  Rp. 250.001 - Rp. 500.000  Tidak Berpenghasilan
```



```

▶ #setting dataset dengan 11 Atribut nilai skor yang tinggi
X = dfBansos[['Pekerjaan KK','Pekerjaan Istri',
              'Penghasilan KK','Penghasilan Istri',
              'Jumlah Tanggungan','Kepemilikan Rumah',
              'Sumber Listrik','Luas Tanah','Luas Bangunan',
              'Sumber Air','MCK']].values #variabel input
y = dfBansos[['Status']].values #varibael target/output

```

```

[ ] # Train test split untuk membagi data training dan testing

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=10)

print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)

```

```

➡ Train set: (107, 11) (107, 1)
   Test set: (27, 11) (27, 1)

```

```

[ ] from sklearn.metrics import accuracy_score
    # Train Naive Bayes classifier
    nb_classifier = GaussianNB()
    nb_classifier.fit(X_train, y_train)

    # Predict on test set
    y_pred = nb_classifier.predict(X_test)

    # Calculate accuracy
    accuracy = accuracy_score(y_test, y_pred)
    print("Accuracy:", accuracy)

```

```

➡ Accuracy: 0.8518518518518519
   /usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1143: DataConversionWarning:
     y = column_or_1d(y, warn=True)

```

```

▶ from sklearn.feature_selection import SelectKBest
  from sklearn.feature_selection import mutual_info_classif
  X = dfBansos.iloc[:,2:13] #independent columns
  y = dfBansos.iloc[:, -1]  # target colum i.e price range

  # apply SelectKBest class to extract

  bestfeatures = SelectKBest(score_func=mutual_info_classif, k=11)
  fit = bestfeatures.fit(X,y)
  dfscores = pd.DataFrame(fit.scores_)
  dfcolumns = pd.DataFrame(X.columns)

  #concat two dataframes for better visualization

  featureScores = pd.concat([dfcolumns,dfscores],axis=1)
  featureScores.columns = ['Atribut','Score'] #naming the dataframe columns
  print(featureScores.nlargest(11,'Score')) #print 10 best features

```

```

➡
   Atribut      Score
2  Penghasilan KK  0.200365
3  Penghasilan Istri  0.197497
0   Pekerjaan KK  0.167354
7    Luas Tanah  0.124624
8   Luas Bangunan  0.083985
10      MCK  0.061642
1   Pekerjaan Istri  0.013110
4   Jumlah Tanggungan  0.000000
5  Kepemilikan Rumah  0.000000
6   Sumber Listrik  0.000000
9    Sumber Air  0.000000

```

```

[ ] #setting dataset dengan nilai skor yang tinggi 7 Atribut
X = dfBansos[['Penghasilan KK','Penghasilan Istri','Pekerjaan KK',
              'Luas Bangunan','Luas Tanah','Sumber Air',
              'MCK']].values #variabel input
y = dfBansos[['Status']].values #varibael target/output

```



```
[ ] #setting dataset dengan nilai skor yang tinggi 7 Atribut
X = dfBansos[['Penghasilan KK','Penghasilan Istri','Pekerjaan KK',
              'Luas Bangunan','Luas Tanah','Sumber Air',
              'MCK']].values #variabel input
y = dfBansos[['Status']].values #varibael target/output
```

```
▶ # Train test split untuk membagi data training dan testing

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=10)

print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
```

```
⇒ Train set: (107, 7) (107, 1)
Test set: (27, 7) (27, 1)
```

```
[ ] from sklearn.metrics import accuracy_score
# Train Naive Bayes classifier
nb_classifier = GaussianNB()
nb_classifier.fit(X_train, y_train)

# Predict on test set
y_pred = nb_classifier.predict(X_test)

# Calculate accuracy
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

```
⇒ Accuracy: 0.9629629629629629
/usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1143: DataConversionWarning:
  y = column_or_1d(y, warn=True)
```

**LAMPIRAN 2**  
**RIWAYAT HIDUP**

Nama : Fahrún Lakadjo  
Jenis Kelamin : Laki-laki  
Tempat Tanggal Lahir : Taluda'a, 01-10-2000  
Pekerjaan : Mahasiswa  
E-mail : [flakadjo@gmail.com](mailto:flakadjo@gmail.com)



**Riwayat Pendidikan :**

1. Tahun 2013, menyelesaikan pendidikan di SDN 1 Bone
2. Tahun 2016, menyelesaikan pendidikan di SMP Negeri 1 Bone
3. Tahun 2019, menyelesaikan pendidikan di SMK Negeri 2 Gorontalo
4. Tahun 2020, diterima menjadi mahasiswa di Perguruan Tinggi Swasta Universitas Ichsan Gorontalo

## LAMPIRAN 3 HASIL TURNITIN



Similarity Report ID: oid:25211:60603413

PAPER NAME

SKRIPSI\_T3120017\_FAHRUN\_LAKADJO.  
pdf

AUTHOR

Fahrin Lakadjo flakadjo@gmail.com

WORD COUNT

11263 Words

CHARACTER COUNT

63188 Characters

PAGE COUNT

58 Pages

FILE SIZE

1.1MB

SUBMISSION DATE

Jun 3, 2024 2:25 PM GMT+8

REPORT DATE

Jun 3, 2024 2:26 PM GMT+8

### ● 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 9% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 1% Submitted Works database

### ● Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 30 words)



100



63	Amran Mulyan	Lainnya	Rp. 250.000 - Rp. 500.000	Tidak Berpenghasilan	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
64	ABDANAN PUTUNUDURO	Lainnya	< Rp. 100.000	< Rp. 250.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
65	Ronaid Umbara	Peg. Swasta	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	>200 M2	>200 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
66	Dodon Lipito	Lainnya	< Rp. 100.000	< Rp. 250.000	3-4 Orang	Tidak Mandiri	PI-N	50-99 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
67	Fajardo Damaynu	Tidak Berpenghasilan	> Rp. 1.000.000	> Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
68	Barni Hapli	Lainnya	< Rp. 100.000	< Rp. 250.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
69	Taufik Lajung	Lainnya	< Rp. 250.000	< Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
70	Taufik Samia	Lainnya	< Rp. 250.000	< Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
71	IBRAHIM SALEH	Tidak Berpenghasilan	Rp. 750.000 - Rp. 1.000.000	Rp. 750.000 - Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
72	SUDONO MANI	Tidak Berpenghasilan	> Rp. 1.000.000	> Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
73	Imet mambli	Peg. Swasta	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
74	Aqon Mabo	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
75	Mudawen Awan	Tidak Berpenghasilan	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
76	ABIN KALATI	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
77	Mohammad Polomilo	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
78	Pradisa Raturio Maki	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
79	IRANADI MANDUBANG	Peg. Swasta	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
80	Jansen B	Lainnya	Rp. 750.000 - Rp. 1.000.000	Rp. 750.000 - Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
81	KANDAN DUAARD	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
82	BAU WIKACON M KARIN	Tidak Berpenghasilan	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
83	Syahrudin Dhar	Peg. Swasta	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
84	VACOD MODATU	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
85	Ran Panto (amamun)	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
86	Turandi Bira	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
87	SABALUDIN DIAWANA	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
88	Abdul Fari Laminu	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
89	YEPRI MUKSIOD	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
90	ITON INGGOLO	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
91	Mahmud Kibana	Tidak Berpenghasilan	> Rp. 1.000.000	> Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
92	Sardin Tasa	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
93	Arpan Nosa	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
94	Harjo Mad	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
95	Warten	Peg. Swasta	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
96	Parin B. Dwendu	Tidak Berpenghasilan	< Rp. 250.000	< Rp. 250.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
97	Small abdu	Lainnya	Rp. 1.000.000	Rp. 1.000.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
98	Satri Huse	Peg. Swasta	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
99	ANWAR NUS	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
100	Wahidun Shudung	Peg. Swasta	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
101	JATAS DUNGGA	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
102	Ku. Kanyor	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
103	Kusman Adin	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
104	Peri Jufi	Peg. Swasta	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
105	ASWAN MOSI	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
106	Andri Ramling	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
107	Dedion Tadiro	Tidak Berpenghasilan	< Rp. 250.000	< Rp. 250.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
108	Arin SA MH	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
109	Imal doli	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
110	SAMIN A. TUA	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
111	Fernando Bira	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
112	Sani Pomonito	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
113	MAKSON A. SAMINIA	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
114	MUSLI LAMON	Lainnya	Rp. 250.000 - Rp. 500.000	Rp. 250.000 - Rp. 500.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
115	Bambang R. Ekmoro	Tidak Berpenghasilan	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
116	Nurdin Dira	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
117	Maslan abas	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
118	IRMAN BADIERA	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
119	VACAN MANGGUA	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
120	Bazari Medin	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
121	BASABIS TOTORE	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
122	KAMDI DUAARD	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
123	Wahid Dalia	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
124	Basim	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
125	MORIN NERTO	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
126	SITI MUHAMMAD PUTRI BUDU	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
127	Nuan Dugano	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
128	SITI NUBAN MAULIT	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima
129	Muh. Fatin Ramadhan Dugano	Lainnya	Rp. 500.000 - Rp. 750.000	Rp. 500.000 - Rp. 750.000	3-4 Orang	Sendiri	PI-N	100-200 M2	25-50 M2	Sumur	Dependikan Sumur Dulu	Tidak Terima



## LAMPIRAN 4

### SURAT TELAH MELAKUKAN PENELITIAN



**PEMERINTAH KABUPATEN BONE BOLANGO**  
**KECAMATAN BONE**  
**DESA MUARA BONE**

*Jln. Trans Sulawesi Lintas Selatan Desa Muara Bone Kode Pos : 96571*

**SURAT KETERANGAN PENELITIAN**

Nomor:140/MB/BN-BB/71/IV/2024

Yang Bertandatangan dibawah ini Kepala Desa Muara Bone Kecamatan Bone, Kabupaten Bone Bolango Provinsi Gorontalo, menerangkan bahwa sesungguhnya :

Nama	: Fahrur Lakadjo
Nim	: T3120017
Fakultas	: Ilmu Komputer
Prodi	: Teknk Informatika
Univeersitas	: Ichsan Gorontalo

Mahasiswa diatas tersebut Benar-benar telah Melakukan observasi dan Penelitian di Desa Muara Bone, Kecamatan Bone, Kabupaten Bone bolango, Provinsi Gorontalo "*Penerapan Information Gain Untuk Seleksi Atribut Pada Klasifikasi Penerima Bantuan Sosial Menggunakan Naive Bayes(Studi Kasus Desa Muara Bone)*"

Adapun Kegiatan Tersebut dilaksanakan Pada bulan Oktober sampai dengan selesai.

Demikian Surat ini dibuat dan digunakan sebagaimana Mestinya

Muara Bone, 25 April 2024

Mengetahui  
Kepala Desa Muara bone





## LAMPIRAN 5 BEBAS PUSTAKA



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI  
**UNIVERSITAS ICHSAN GORONTALO**  
FAKULTAS ILMU KOMPUTER  
UPT. PERPUSTAKAAN FAKULTAS  
SK. MENDIKNAS RI NO. 84/D/0/2001  
Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

### **SURAT KETERANGAN BEBAS PUSTAKA**

No : 015/Perpustakaan-Fikom/V/2024

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Fahrur Lakadjo  
No. Induk : T3120017  
No. Anggota : M202420

Terhitung mulai hari, tanggal : Jumat, 31 Mei 2024, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di gunakan sebagaimana mestinya.



Gorontalo, 31 Mei 2024

Mengetahui,  
Kepala Perpustakaan

**Apriyanto Alhamad, M.Kom**

**NIDN : 0924048601**



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

**UNIVERSITAS ICHSAN GORONTALO**

**FAKULTAS ILMU KOMPUTER**

**SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001**

Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

**SURAT REKOMENDASI BEBAS PLAGIASI**

No. 105/FIKOM-UIG/R/VI/2024

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom  
NIDN : 0928028101  
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

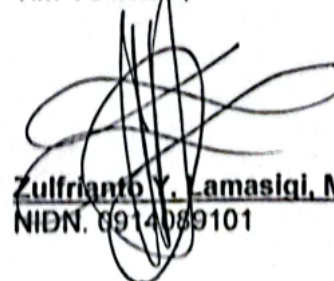
Nama Mahasiswa : Fahrur Lakadjo  
NIM : T3120017  
Program Studi : Teknik Informatika (S1)  
Fakultas : Fakultas Ilmu Komputer  
Judul Skripsi : Penerapan Information Gain Untuk Seleksi Atribut  
Pada Klasifikasi Penerima Bantuan Sosial  
Menggunakan Naive Bayes

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **9%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui  
Dekan,  
  
**Irvan Abraham Salihi, M.Kom**  
NIDN. 0928028101

Gorontalo, 4 Juni 2024  
Tim Verifikasi,

  
**Zulfrianto Y. Lamasigi, M.Kom**  
NIDN. 0914089101

Terlampir :  
Hasil Pengecekan Turnitin