

IMPLEMENTASI METODE *NAÏVE BAYES* UNTUK KLASIFIKASI STATUS EKONOMI MASYARAKAT

(Studi kasus : Desa Dambalo)

Oleh

RAFLIN MUSA

T3117233

SKRIPSI

Untuk memenuhi salah satu syarat ujian

Guna memperoleh gelar sarjana program studi teknik informatika,

Ini telah disetujui oleh tim pembimbing

Gorontalo, 8 Desember 2022



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2022**

**PERSETUJUAN SKRIPSI
IMPLEMENTASI METODE *NAÏVE BAYES* UNTUK
KLASIFIKASI STATUS EKONOMI MASYARAKAT**

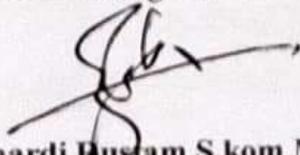
(Studi kasus : Desa Dambalo)

Oleh
RAFLIN MUSA
T3117233

SKRIPSI

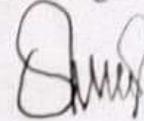
Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar sarjana program studi teknik informatika,
Ini telah disetujui oleh tim pembimbing
Gorontalo, 28 Desember 2022

Pembimbing Utama



Suhardi Rustam S.kom M.kom
NIDN : 0915088403

Pembimbing Pendamping



Sumarni, S.kom M.kom
NIDN : 0926018604

PENGESAHAN SKRIPSI

IMPLEMENTASI METODE *NAÏVE BAYES* UNTUK
KLASIFIKASI STATUS EKONOMI MASYARAKAT

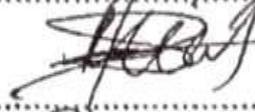
(Studi kasus : Desa Dambalo)

Oleh
RAFLIN MUSA
T3117233

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

1. Ketua Penguji
Muis Nanja, M.Kom
2. Anggota
Sunarto Taliki, M.Kom
3. Anggota
Sarlis Mooduto, M.Kom
4. Anggota
Suhardi Rustam, M.Kom
5. Anggota
Sumarni, M.Kom

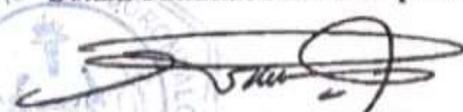

.....

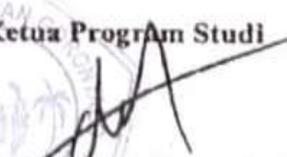
.....

.....

.....

.....

Mengetahui


Dekan Fakultas Ilmu Komputer

Irvan Abraham Salihi, M.Kom
NIDN 0928028101


Ketua Program Studi

Sudirman S. Panna M.Kom
NIDN 0924038205

PERNYATAAN SKRPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelas akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun diperguruan 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 ebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat pem=nyimpangan 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, 28 Desember 2022

Yang membuat pernyataan



Rafin Musa

ABSTRACT

RAFLIN MUSA. T3117233. THE IMPLEMENTATION OF THE NAÏVE BAYES METHOD FOR COMMUNITY ECONOMIC STATUS CLASSIFICATION (CASE STUDY: DAMBALO VILLAGE, TOMILITO SUB- DISTRICT, NORTH GORONTALO DISTRICT)

The research aims to find out: 1) the process of analyzing data grouping the economic status of the community using the Naïve Bayes method, 2) the implementation of Naïve Bayes in the economic status grouping of the Dambalo Village community in Tomilito Subdistrict, North Gorontalo District. This research employs a qualitative approach, and the method used is the observation which is done by direct observation at the research site. The object of research is the implementation of the Naïve Bayes method for the classification of the economic status of the community. The research conducted spends 7 months starting from April 2022 to November 2022 at the Dambalo Village Office, Tomiito Subdistrict, North Gorontalo District. Based on the classification results, three (3) labels are obtained, namely Rich, Middle, and Poor. The results of the discussion in this study begin with the stages of interviews and data documentation then proceed with the calculation of the algorithm used. From the classification results, 100% accuracy is obtained on Training Data, namely 45 Data. The results of the classification of test data totaling 5 data obtained are D1, D2, D3, and D5 have poor economic status and D4 has a medium economic status. The Naïve Bayes method can help the government, especially the Dambalo Village government, in maximizing the processing of community economic status data.

Keywords: data mining, Naïve Bayes, classification, economic status

ABSTRAK

Raflin Musa, T3117233. Implementasi Metode *Naïve Bayes* Untuk Klasifikasi Status Ekonomi Masyarakat (Studi kasus : Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara).

Penelitian bertujuan 1). Untuk mengetahui proses analisis data pengelompokan status ekonomi masyarakat menggunakan metode *Naïve Bayes*. 2). Untuk mengetahui penerapan *Naïve Bayes* pada pengelompokan status ekonomi masyarakat Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara. Pada penelitian ini menggunakan pendekatan kualitatif dengan metode yang digunakan yaitu observasi yang dilakukan dengan pengamatan secara langsung di tempat penelitian. Maka yang akan menjadi objek penelitian adalah implementasi metode *naïve bayes* untuk klasifikasi status ekonomi masyarakat. Penelitian ini dilaksanakan dalam waktu 7 bulan terhitung mulai dari bulan April 2022 sampai November 2022 bertempat di Kantor Desa Dambalo Kecamatan Tomiito Kabupaten Gorontalo Utara. Dari hasil klasifikasi diperoleh 3 Label yakni Kaya, Menengah, Miskin. Hasil pembahasan dalam penelitian ini dimulai dari tahapan wawancara dan dokumentasi data kemudian dilanjutkan dengan perhitungan algoritma yang digunakan. Dari hasil klasifikasi diperoleh akurasi 100% pada Data Training sebanyak 45 Data. Hasil klasifikasi Data Uji sejumlah 5 Data diperoleh D1, D2, D3, D5 memiliki status ekonomi Miskin dan D4 memiliki status ekonomi Menengah. Maka metode *Naïve Bayes* dapat membantu pemerintah khususnya pemerintah Desa Dambalo dalam memaksimalkan pengolahan data status ekonomi masyarakat.

Kata Kunci : Data Mining, *Naïve Bayes*, Klasifikasi, Status Ekonomi

KATA PENGANTAR

Alhamdulillahirobbil'alamiin, penulis dapat menyelesaikan skripsi ini dengan judul: ***“IMPLEMENTASI METODE NAÏVE BAYES UNTUK KLASIFIKASI STATUS EKONOMI MASYARAKAT”***, untuk memenuhi salah satu syarat memperoleh gelar Sarjana Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa penulisan 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 Abdulssamad M.Si selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar Latjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi, M.kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman S. Melangi, S.Kom M.Kom, selaku pembantu Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, S.Kom, M.Kom, selaku pembantu Dekan II Bidang Adminstrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Bapak Sudirman S. Panna, S.Kom M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Bapak Suhardi Rustam, S.kom M.kom, selaku Pembimbing Utama yang telah membimbing penulis selama menyusun usulan penelitian;
8. Ibu Sumarni, S.kom M.kom, selaku Pembimbing Pendamping yang telah membimbing penulis selama menyusun usulan penelitian;

9. Aparat dan Ayahanda Kepala Desa Dambalo yang telah membantu penulis dalam pengumpulan data di lapangan;
10. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
11. Orang tua yang tercinta, atas segala kasih sayang, jerih payah dan doa restunya dan membersarkan serta mendidik penulis;
12. Kakak - kakak yang selalu memotivasi dan menyemangati penulis;
13. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
14. Kepada semua pihak yang ikut membantu penyelesaian skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Penulis menyadari sepenuhnya bahawa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya ktirik dan saran yang konstruktif. Akhirnya, penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua, Aamiin.

Gorontalo, 8 Desember 2022

Raflin Musa

DAFTAR ISI

HALAMAN SAMPUL

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
HALAMAN PERNYATAAN.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii

BAB I PENDAHULUAN

1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.5.1 Manfaat Teoritis	4
1.5.2 Manfaat Praktis	4

BAB II TINJAUAN PUSTAKA

2.1 Tinjauan Studi	5
2.2 Tinjauan Pustaka	6

2.2.1 Status Ekonomi Masyarakat.....	6
2.2.1 Data Mining.....	7
2.2.3 Klasifikasi.....	9
2.2.4 Metode Naive Bayes	9
2.2.5 Penerapan Algoritma Naive Bayes	10
2.2.6 Rapid Miner.....	12
2.3 Kerangka Pikir.....	14

BAB III METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Waktu dan Lokasi Penelitian.....	15
3.2 Pengumpulan Data	15
3.3 Pemodelan Naive Bayes	16
3.4 Pengujian dan Validasi	17

BAB IV HASIL PENELITIAN

4.1 Hasil pengumpulan Data.....	19
4.2 Penjelasan Algoritma	22
4.3 Perhitungan Naive Bayes	22
4.4 Perhitungan Excel	26
4.4.1 Probabilitas Tiap Label	26
4.4.2 Probabilitas Kriteria Tempat Tinggal	26
4.4.3 Probabilitas Kriteria Jenis Tempat Tinggal	27
4.4.4 Probabilitas Kriteria Pekerjaan	27
4.4.5 Probabilitas Kriteria Umur.....	28
4.4.6 Probabilitas Kriteria Tanggungan	29
4.4.7 Hasil Perhitungan Data Uji	29

BAB V PEMBAHASAN PENELITIAN

5.1 Halaman Import Data.....	30
------------------------------	----

5.2 Tampilan Select Location Data.....	30
5.3 Review Tampilan Data	31
5.4 Input Algoritma Naive Bayes	31
5.5 Input <i>Apply Model</i> dan <i>Performance</i>	32
5.6 Hubungkan semua operator	32
5.7 Proses Algoritma Naive Bayes	33
5.8 Hasil perhitungan data uji	34
BAB VI PENUTUP	
6.1 Kesimpulan	35
6.2 Saran	35
DAFTAR PUSTAKA	36
LAMPIRAN.....	38

DAFTAR GAMBAR

Gambar 2.1 Proses KDD	7
Gambar 2.2 Alur Metode Naïve Bayes	9
Gambar 3.1 Pemodelan.....	16
Gambar 5.1 Halaman Import Data	27
Gambar 5.2 Pilih Data Training dan DataUji	28
Gambar 5.3 Review Tampilan Data	28
Gambar 5.4 Input Algoritma Naive Bayes	29
Gambar 5.5 <i>Apply Model</i> dan <i>Performance</i>	29
Gambar 5.6 Hubungkan semua operator	30
Gambar 5.7 Hasil Akurasi Data Training	30
Gambar 5.8 Hasil Data Uji	31

DAFTAR TABEL

Tabel 1.1 Sampel Data	3
Tabel 2.1 Penelitian Terkait	5
Tabel 3.1 Atribut Data.....	16
Tabel 4.1 Probilitas Tiap Label	24
Tabel 4.2 Probilitas Kriteria Tempat Tinggal.....	24
Tabel 4.3 Probilitas Kriteria Jenis Tempat Tinggal	25
Tabel 4.4 Probilitas Kriteria Pekerjaan	25
Tabel 4.5 Probilitas Kriteria Umur.....	26
Tabel 4.6 Probilitas Kriteria Tanggungan	26
Tabel 4.7 Perhitungan Data Uji	26

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembangunan desa dilakuka guna memperkuat dasar perekonomian bangsa, memberantas kemiskinan dan pengurangi kesenjangan perkembangan antar wilayah, sebagai jawaban bagi perubahan sosial, desa sebagai dasar perubahan. Pembangunan desa sejarang ini mengalami perubahan yang cukup signifikan dalam konsep maupun prosesnya. Konsep pembangunan tidak hanya sebatas pada wilayah agraris dan infrastruktur dasar tetapi mengarah pada pengembangan Teknologi Informasi dan Komunikasi (TIK). Pengguna teknologi informasi dan komunikasi memberikan efek positif adalah menjadikan suatu pekerjaan menjadi lebih efektif dan efisien [1].

Dalam Pada suatu wilayah terdapat kumpulan-kumpulan masyarakat yang mempunyai status ekonomi yang berbeda-beda. Perlunya mengetahui perbedaan status ekonomi masyarakat agar nantinya jika ada bantuan dari pemerintah untuk masyarakat kurang mampu, pendataan warga kurang mampu, dan lainnnya diharapkan dapat tepat sasaran sehingga akan benar-benar membantu masyarakat yang membutuhkan.

Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara merupakan salah satu kelurahan yang memiliki peranan penting untuk mensejahterakan masyarakat salah satunya dengan bantuan pemerintah. Saat ini pendataan status ekonomi masyarakat di Desa tersebut masih kurang objektif sebab tidak sinkronnya pendataan yang dilakukan oleh petugas terkait dengan data yang sebenarnya serta proses klasifikainya hanya dilakukan ketika diperlukan saja, misalnya ketika ada bantuan dari pihak pemerintah. Hal dinilai kurang efektif karena tidak sesuai kriteria yang ditetapkan pemerintah. Jawaban dari masalah tersebut adalah diperlukan untuk dapat mengolah data masyarakat agar memperoleh data status ekonomi masyarakat yang sesuai. Salah satu metode yang dapat digunakan dalam pengolahan data adalah *Data mining*.

Data Mining adalah suatu pekerjaan ataupun kegiatan guna mengumpulkan data yang memiliki berukuran besar lalu mengekstraksi data tersebut menjadi informasi – Informasi yang nanti bisa dipergunakan. Klasifikasi adalah kegiatan penemuan model atau fungsi yang memberikan gambaran dan memisahkan kelas data atau konsep yang bertujuan agar bisa dipergunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Terdapat beberapa algoritma klasifikasi seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes Classifier* (NBC). Salah satu metode klasifikasi yang banyak digunakan secara luas adalah *Naïve Bayes Cluster* (NBY) adalah sebuah kegiatan klasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas ataupun kemungkinan dengan menjumlahkan frekuensi dan kombinasi nilai dari data training yang diberikan [2].

Naive bayes didasari pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu [3]. Keuntungan penggunaan *naive bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [4].

Pada penelitian sebelumnya metode *Naïve bayes* juga pakai dalam memprediksi berbagai macam penyakit Dermatologi yang diabaikan tapi bahkan dapat menyebabkan kematian dimana metode *Naïve bayes* digunakan guna mengenal pola data agar mengetahui kemungkinan penyakit [5]. Metode *Naive bayes* juga diketahui berpotensi baik dalam kegiatan klasifikasi dokumen dibandingkan metode klasifikasi yang lain dalam hal akurasi dan efisiensi komputasi [6].

Proses klasifikasi status ekonomi masyarakat dilakukan dengan berdasarkan variabel umur, jumlah tanggungan, tempat tinggal, jenis tempat tinggal dan pekerjaan. Kemudian akan menghasilkan output kaya, mengah dan miskin sebagai label probabilitas. Dataset yang diperoleh dari tempat peneltian dapat dilihat pada tabel berikut.

Berikut merupakan sampel data yang diperoleh peneliti dari tempat penelitian dapat dilihat pada Tabel 1.1

Tabel 1.1 Sampel Data

NO	NAMA	UMUR	JUMLAH TANGGUNGAN	TEMPAT TINGGAL	JENIS TEMPAT TINGGAL	PEKERJAAN	LABEL
1	HASAN ABUNA	69	3	SENDIRI	PERMANEN	PETANI	MSKIN
2	YOFAN ABD GIAS TANANI	27	2	ORANG TUA	PERMANEN	KARYAWAN SWASTA	KAYA
3	FELDY YUSUF	41	3	SENDIRI	PERMANEN	PETANI	MENENGAH
4	HUSIN AHMAD	49	1	SENDIRI	KAYU	PETANI	MSKIN
5	FARDAN K LABANGA	47	2	SENDIRI	PERMANEN	PNS	KAYA
6	BASRIN GOU	35	3	ORANG TUA	PERMANEN	PNS	KAYA
7	DARWIN ABAY	46	6	SENDIRI	PERMANEN	NELAYAN	MSKIN
8	RIYANTO RAJAK	36	3	SENDIRI	KAYU	BURUH BANGUNAN	MSKIN
9	RAIS TANAIYO	29	3	SENDIRI	KAYU	PETANI	MENENGAH
10	ROMAN TANAIYO	42	4	SENDIRI	KAYU	PETANI	MSKIN

(Sumber : Desa Dambalo, 2021)

Mengacu pada data dan proses kerja yang sudah dijelaskan diatas maka perlu adanya suatu analisis untuk mengelompokan status ekonomi masyarakat dengan metode *Naïve Bayes*, maka penulis termotivasi untuk mengangkat penelitian dengan judul *“Implementasi Metode Naïve Bayes Untuk Klasifikasi Status Ekonomi Masyarakat (Studi Kasus Desa Dambalo)”* diharapkan dari penelitian ini dapat memberikan kontribusi yang dapat digunakan oleh perangkat desa dalam pengambilan keputusan.

1.2 Identifikasi Masalah

1. Lemahnya proses klasifikasi status ekonomi masyarakat karena masih dilakukan secara manual serta berdasarkan perkiraan atau pemikiran manusia.
2. Belum adanya proses analisa dalam klasifikasi status ekonomi masyarakat.

1.3 Rumusan Masalah

1. Bagaimana proses analisis data klasifikasi status ekonomi masyarakat menggunakan metode *Naïve Bayes*?

2. Bagaimana menerapkan *Naïve Bayes* pada klasifikasi status ekonomi masyarakat Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara ?

1.4 Tujuan penelitian

1. Untuk mengetahui proses analisis data klasifikasi status ekonomi masyarakat menggunakan metode *Naïve Bayes*.
2. Untuk mengetahui penerapan *Naïve Bayes* pada klasifikasi status ekonomi masyarakat Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara.

1.5 Manfaat penelitian

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan bisa memberikan masukan bagi perkembangan ilmu pengetahuan dan teknologi, khususnya pada bidang ilmu komputer, yaitu berupa analisis pengelompokan status ekonomi masyarakat menggunakan metode *Naïve Bayes*

1.5.2 Manfaat Praktis

Sebagai solusi bagi perangkat desa dalam memberikan gambaran saat pengambilan keputusan.

BAB II

LANDASAN TEORI

2.1. Tinjauan Studi

Adapun penelitian terkait tentang Implementasi metode *Naive Bayes* untuk mengelompokkan status ekonomi masyarakat, sebagai berikut :

Table 2.1 Penelitian terkait

NO	PENELITI	JUDUL	METODE	HASIL
1	Mohammad Guntura, Julius Santonyb, Yuhandric	Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko	Naïve Bayes	Hasil klasifikasi menggunakan algoritma Naïve Bayes Classifier di uji dengan aplikasi RapidMiner. Hasil klasifikasi dari 16 data yang diuji dengan RapidMiner memiliki nilai akurasi sebesar 75% dan bisa dikatakan algoritma Naïve Bayes Classifier bisa klasifikasi harga emas dengan baik. 5

2	Asrul Ashari Muin	Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi) Karyawan Baru	Naïve Bayes	Naïve Bayes bisa mengerjakan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode Naïve Bayes Clasification
3	Monica Yoshe Titimeidara a, Wiwien Hadikurniawati b	Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita	Naïve Bayes	Berdasarkan penelitian, Metode Naive Bayes Classifier dapat digunakan untuk klasifikasi status gizi stunting pada balita berdasarkan jenis kelamin, umur, berat badan, tinggi badan, status miskin dan kategori status gizi. Hasil akurasi dari Metode Naive Bayes

				Classifier dalam melakukan klasifikasi status gizi stunting pada balita adalah sebesar 88 % dari 300 data yang dibagi menjadi 2 yaitu 275 data sebagai data latih dan 25 data sebagai data uji
--	--	--	--	--

2.2. Tinjauan Pustaka

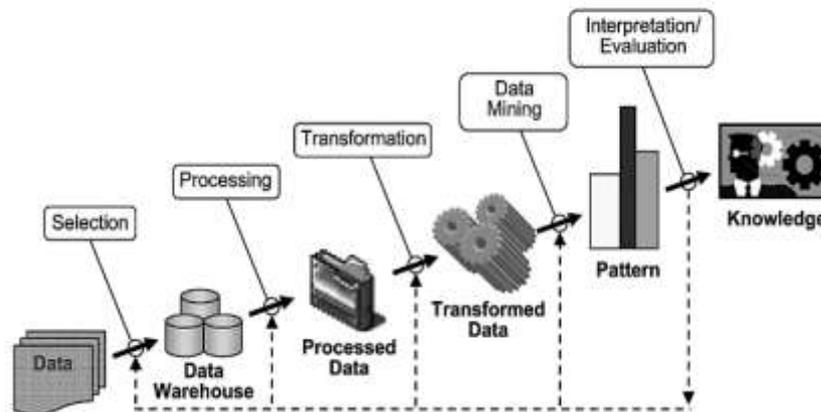
2.2.1 Status Ekonomi Masyarakat

Status sosial ekonomi secara khusus merupakan posisi yang ditempati individu atau keluarga yang berkenaan dengan standar normatif yang umum berlaku tentang kepemilikan kultural, pendapatan efektif, pemilikan barang dan partisipasi dalam aktifitas kelompok dari komunitasnya. Soekanto mengemukakan faktor yang mempengaruhi status sosial ekonomi yakni; pekerjaan, pendidikan, pendapatan, jumlah tanggungan, pemilikan, dan jenis tempat tinggal [2].

2.2.2 Data Mining

Data mining yaitu proses sering di gunakan teknik statistisk mtematika,machine dan kecerdana buatan dapat mengekstrak dan mengindentifikasi ilmu pengetahuan terkait dari *database*. Proses ini bertujuan untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki [3]. Definisi lainnya Data Mining merupakan proses yang meliputi kumpulan, penggunaan data history guna mencari kesesuaian, pola atau hubungan dalam data mempunyai ukuran besar [7]. Dari berbagai pengertian di atas dapat disimpulkan bahwa Data Mining adalah kegiatan ataupun proses untuk menyatukan data yang besar setelah itu menyederhanakan data itu menjadi informasi – informasi yang nanti bisa dipakai.

Berikut adalah proses pengambilan data atau biasa disebut dengan KDD (*Knowledge Discovery in Databases*) :



Gambar 2.1 Proses KDD
(Sumber : Pratama, 2010)

Berikut adalah penjelasan dari proses KDD :

1) *Cleansing*

Adalah proses pembersihan data yang tidak lengkap dan mengandung error dari penyimpanan data.

2) *Integration*

Proses integrasi adalah proses mengkombinasikan jika ada data yang berulang.

3) *Selection*

Proses pemilihan atau seleksi data dari sekumpulan data

4) *Transformation*

Proses transformasi adalah coding pada data yang telah dipilih sehingga sesuai dengan proses data mining.

5) *Data Mining*

Proses Data Mining adalah proses pencarian pola, informasi dari data yang terpilih dengan menggunakan metode dan algoritma tertentu.

6) *Pattern Evolution*

Proses dimana pola yang telah didapatkan perlu ditampilkan dan diidentifikasi apakah sudah sesuai dengan hipotesis.

7) *Knowledge Presentation*

Merupakan proses akhir dimana menjadi tahap visualisasi kepada user agar user mengerti dan menginterpretasikan hasil.

2.2.3 Klasifikasi

Klasifikasi merupakan sebuah kegiatan untuk memperoleh model yang memberikan gambaran dan membedakan label dari konsep data. Model ini merupakan dasar analisa satu set data uji, pemodelan ini dipakai guna memprediksi label kelas objek variabel labelnya belum diketahui sebelumnya [4]. Klasifikasi data terdiri dari 2 langkah proses. Pertama merupakan fase belajar atau learning (*fase training*), yakni algoritma klasifikasi dibuat guna menganalisis data uji kemudian direpresentasikan ke bentuk rules klasifikasi. Proses kedua merupakan klasifikasi, yakni data training digunakan untuk memperkirakan akurasi dari rule klasifikasi

2.2.4 Metode *Naive Bayes*

Naive Bayes merupakan satu dari beberapa algoritma pembelajaran induktif yang sangat efektif dan efisien untuk mesin belajar dan data mining. Performance *naive bayes* yang kompetitif pada proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian *naive bayes* cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan [8].

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [9]. Definisi lain

mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [10]

2.2.5 Penerapan Algoritma *Naive Bayes*

Teori keputusan bayes merupakan rumus statistik yang fundamental dalam memperkenalkan pola (*pattern recognition*), kedekatan ini didasarkan pada kuantifikasi *trade-off* antara beberapa keputusan klasifikasi dalam menggunakan kemungkinan dan ongkos yang di timbul pada keputusan tersebut. Selain itu Bayesian classification juga dapat memprediksi probabilitas keanggotaan suatu *Class* pada *teorema bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [11]. Teorema Bayes memiliki bentuk umum sebagai berikut :

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

Keterangan:

X = Data dengan Class yang belum diketahui

H = Hipotesis data X merupakan suatu Class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

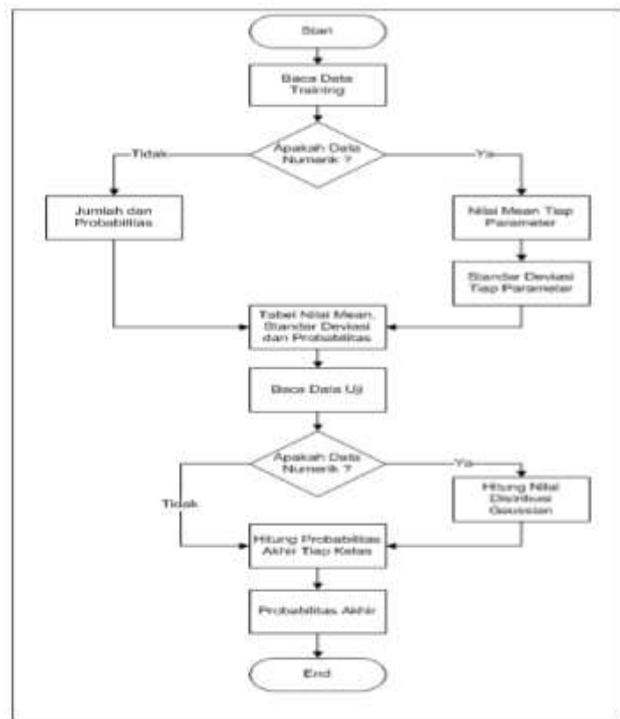
$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

Di mana Variabel X menjabarkan kelas, sementara variabel menjabarkan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Rumus tersebut menjabarkan bahwa peluang masuknya sampel data karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel data tersebut, seringkali disebut prior), dikali dengan peluang kemunculan

karakteristik-karakteristik sampel data pada kelas C (disebut juga likelihood), dibagi dengan oportunitas kemunculan karakteristik-karakteristik sampel secara keseluruhan (disebut juga evidence).



Gambar 2.2 Alur Metode *Naïve Bayes*

Alur dari metode *Naïve Bayes* dapat dilihat pada penjelasan berikut.

1. Baca data training
2. Hitung Jumlah dan probabilitas
 - a) Jika terdapat data numerik, maka temukan nilai mean dan standar deviasi dari masing-masing parameter yang menggambarkan data angka.
Rumus yang digunakan untuk menghitung nilai rata – rata hitung (mean) dapat dilihat sebagai berikut :

Rumus menghitung mean

$$\mu = \sum_{i=1}^n x_i$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 \dots + x_n}{n}$$

dimana :

μ : rata – rata hitung (mean)

x_i : nilai sample ke - i

n : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat dibawah ini:

Rumus simpangan baku

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

Dimana :

σ : standar deviasi

x_i : nilai x ke - i

μ : rata-rata hitung

n : jumlah sampel

- b) Jika tidak, hitung nilai probabilitas tiap kategori yang sama, dengan cara jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut untuk menemukan nilai probabilistik.

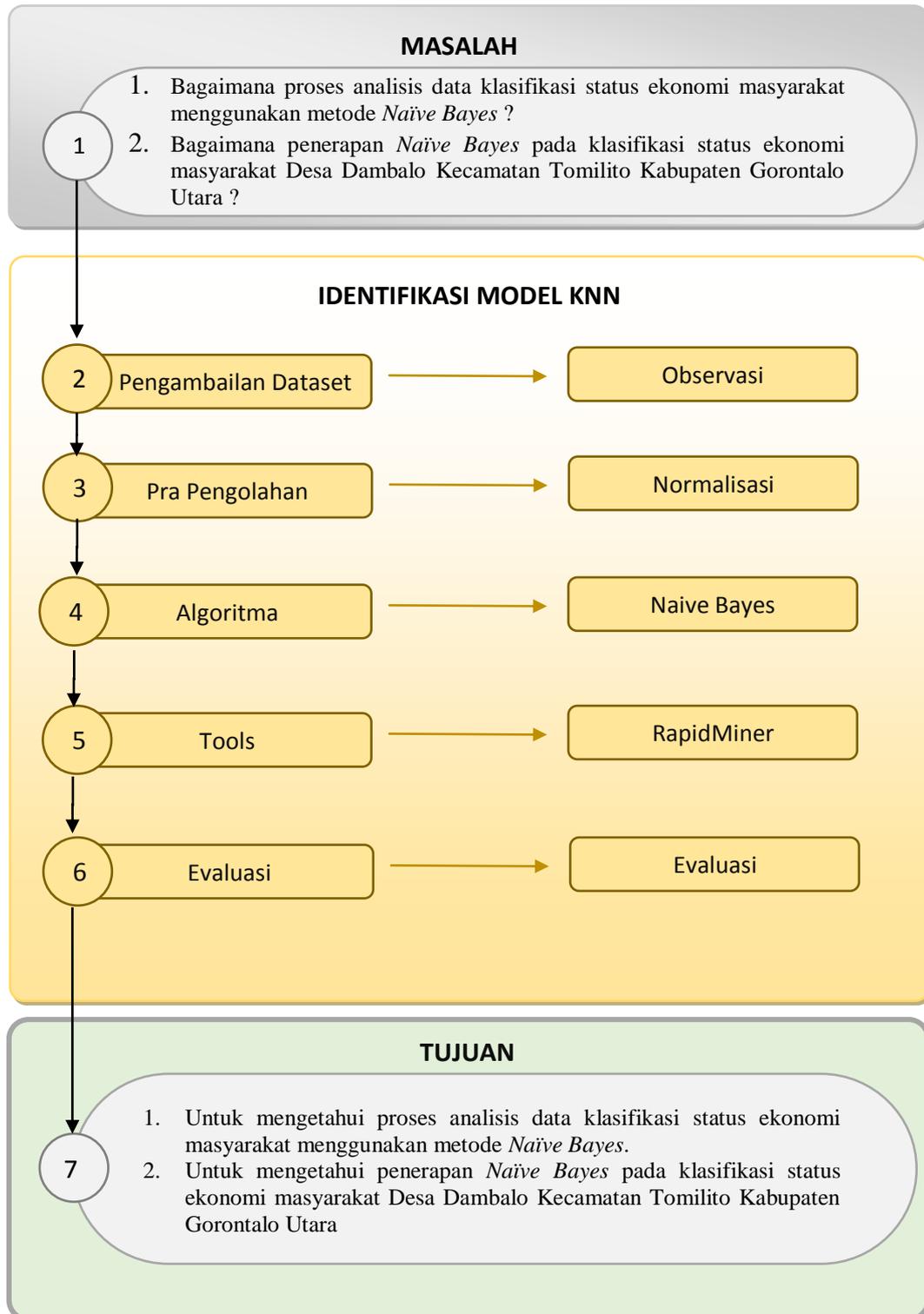
2.2.6 Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner merupakan software yang

berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik[5].

Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan Rapid Miner pun dapat ditampilkan secara visual dengan grafik. Menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining Brilian Rahmat C.T.I. [12].

2.3 Kerangka pikir



BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Waktu dan Lokasi Penelitian

Penelitian ini merupakan penelitian terapan jika dipandang dari tingkat penerapan karena penelitian ini berfokus penerapannya untuk memberikan solusi atas permasalahan secara praktis. Dipandang dari informasi yang diolah, maka penelitian ini merupakan penelitian kuantitatif. Dipandang dari perlakuan terhadap data, maka penelitian ini merupakan penelitian konfirmatori. Penelitian ini menggunakan metode studi kasus dengan demikian jenis penelitian ini adalah deskriptif.

Berdasarkan pada latar belakang dan kerangka fikir yang telah di uraikan maka yang menjadi obyek penelitian adalah Implementasi metode Naïve Bayes untuk mengelompokkan status ekonomi masyarakat Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara. Penelitian ini dimulai dari April 2021 sampai dengan September 2021 yang berlokasi di Kantor Desa Dambalo.

3.2 Pengumpulan Data

Metode pengumpulan data yang digunakan untuk mendapatkan data dan informasi digunakan 2 (dua) jenis data yaitu data primer dan data sekunder sebagai berikut :

1. Penelitian Data Primer (Lapangan)

- a. Metode Observasi

Dalam hal ini yang akan dilakukan adalah melihat serta mempelajari permasalahan yang ada dilapangan yang erat kaitannya dengan objek yang diteliti.

- b. Wawancara

Metode ini digunakan dengan mengajukan beberapa pertanyaan kepada Aparat Desa Dambalo untuk mengetahui permasalahan di lapangan.

2. Penelitian Data Sekunder (Kepustakaan)

Metode kepustakaan diperlukan untuk mendapatkan data sekunder dengan tujuan melengkapi data primer. Data sekunder didapatkan dari pengkajian kepustakaan yang berisi dasar-dasar teori. Metode kepustakaan digunakan oleh analis sistem dengan cara mengambil contoh dokumen-dokumen yang berhubungan dengan materi penelitian. Selain itu, analis sistem mencari data mengenai hal-hal atau variabel yang berupa catatan, buku, majalah, dan sebagainya yang berhubungan dengan penelitian.

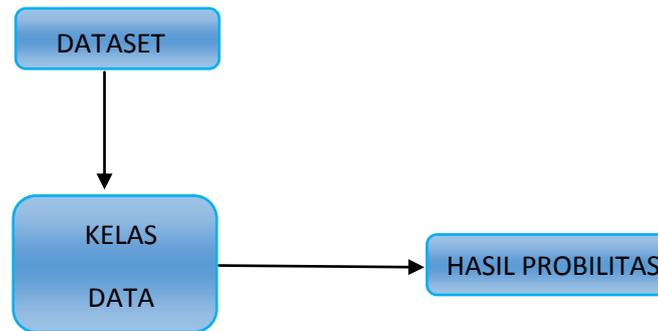
Adapun variabel atau atribut dengan tipe datanya masing- masing ditunjukkan pada tabel berikut ini :

Tabel 3.1 Atribut Data

NO	NAME	TYPE	KET
1	Nama	Binominal	Variabel Input
2	Umur	Nominal	Variabel Input
3	Jumlah Tanggungan	Nominal	Variabel Input
4	Tempat Tinggal	Binominal	Variabel Input
5	Jenis Tempat Tinggal	Binominal	Variabel Input
6	Pekerjaan	Binominal	Variabel Input
7	Kaya	Binominal	Label
8	Menengah	Binominal	Label
9	Miskin	Binominal	Label

3.3 Pemodelan Naïve Bayes

Sistem yang diusulkan dapat digambarkan menggunakan flowchart sebagai berikut :



Gambar 3.1 Pemodelan

3.4 Pra Pengolahan

Sebelum data dikelola pada awalnya perlu dilakukan pembersihan data dengan menghapus data yang tidak relevan. Hal ini perlu dilakukan karena data yang telah diperoleh dari tempat penelitian memiliki isi yang belum sempurna, baik dari data yang hilang, tidak valid atau bisa menjadi data yang salah dalam pengetikan dan data yang tidak relevan lebih baiknya digunakan.

3.5 Evaluasi

Evaluasi adalah bertujuan untuk mengetahui hasil kinerja dari metode yang telah digunakan, evaluasi ini dilakukan untuk menatah hasil kriteria- kriteria dari data untuk mengelompokan status ekonomi masyarakat menggunakan metode Naïve Bayes.

3.6 Pengujian Dan Validasi

Proses pengujian menggunakan tool rapidminer dan melihat data apakah sesuai dengan hasil yang diperoleh melalui tool tersebut. Sedangkan validasi metode dan algoritma Naïve Bayes dilakukan dengan mengukur hasil accuracy, percision dan recall dan dapat dihitung dengan menggunakan Confusion Matrix sebagai berikut : Nilai accuracy dihitung dengan cara menjumlah data benar yang

bernilai positif (*True Positive*) ditambah dengan nilai Negatif (*True Negative*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*), Negatif (*True Negative*) dan ditambah dengan data salah yang bernilai positif (*False Positif*), Negatif (*False Negative*). Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengambilan Data

Data peneliti peroleh dengan cara wawancara dan observasi lokasi penelitian yang dilakukan di Kantor Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara. Setelah dilakukan wawancara dan observasi diperoleh masalah yang ada di lokasi penelitian adalah masih adanya kesulitan untuk membagi status ekonomi masyarakat Desa Dambalo. Berikut data yang diperoleh dari lokasi penelitian :

Tabel 4.1 Hasil Pengumpulan Data

No	Nama	Umur	Jumlah Tanggungan	Tempat Tinggal	Jenis Tempat Tinggal	Pekerjaan	Label
1	Hasan Abuna	69	3	Sendiri	Permanen	Petani	Miskin
2	Yofan Abd Gias Tanani	27	2	Orang Tua	Permanen	Karyawan Swasta	Kaya
3	Feldy Yusuf	41	3	Sendiri	Permanen	Petani	Menengah
4	Husin Ahmad	49	1	Sendiri	Kayu	Petani	Miskin
5	Fardan K Labanga	47	2	Sendiri	Permanen	Pns	Kaya
6	Basrin Gou	35	3	Orang Tua	Permanen	Pns	Kaya
7	Darwin Abay	46	6	Sendiri	Permanen	Nelayan	Miskin
8	Riyanto Rajak	36	3	Sendiri	Kayu	Buruh Bangunan	Miskin
9	Rais Tanaiyo	29	3	Sendiri	Kayu	Petani	Menengah

10	Roman Tanaiyo	42	4	Sendiri	Kayu	Petani	Miskin
11	Moh. Hidayat Rasyid	24	2	Orang Tua	Permanen	Buruh Bangunan	Menengah
12	Azis Ngila	43	5	Sendiri	Permanen	Petani	Menengah
13	Tamrin Tahir	67	1	Sendiri	Kayu	Nelayan	Miskin
14	Agus Donto	34	3	Sendiri	Permanen	Nelayan	Menengah
15	Jon Donto	52	5	Sendiri	Permanen	Nelayan	Miskin
16	Rafik A. Kau	29	2	Bersama Orang Tua	Permanen	Pns	Kaya
17	Azis R Tue	48	3	Sendiri	Permanen	Petani	Menengah
18	Saleh Tueno	70	1	Sendiri	Permanen	Petani	Menengah
19	Iwan Suna	47	2	Sendiri	Permanen	Nelayan	Menengah
20	Lamrin Mooduto	43	4	Sendiri	Permanen	Petani	Menengah
21	Rukmin Tanaiyo	62	1	Sendiri	Kayu	Irt	Miskin
22	Sumarni Hipi	42	2	Sendiri	Permanen	Irt	Menengah
23	Monira Umar	64	2	Sendiri	Permanen	Irt	Miskin
24	Sukrin Ishak	36	3	Orang Tua	Permanen	Perangkat Desa	Menengah
25	Maryam Baingo	81	0	Sendiri	Permanen	Irt	Menengah
26	Romen Lingude	35	3	Sendiri	Kayu	Petani	Menengah
27	Asjura Kudi	56	3	Sendiri	Kayu	Petani	Miskin
28	Rustam Hatu Djailani	62	1	Sendiri	Kayu	Petani	Miskin
29	Drs. Hj Supratman Nina	57	2	Sendiri	Permanen	Pns	Kaya
30	Arpan W Nur	28	2	Orang Tua	Permanen	Petani	Menengah

31	Hani Ahmad	82	1	Sendiri	Permanen	Irt	Miskin
32	Riskiyanto Tanaiyo	23	2	Orang Tua	Permanen	Buruh Bangunan	Menengah
33	Suleman Tomayahu	21	2	Orang Tua	Permanen	Petani	Menengah
34	Yuyu Donto	58	1	Sendiri	Permanen	Petani	Menengah
35	Marice Datau	85	1	Sendiri	Permanen	Irt	Miskin
36	Abdul Azis	36	3	Sendiri	Permanen	Buruh Bangunan	Menengah
37	Sarli Tanaiyo	27	2	Orang Tua	Permanen	Petani	Menengah
38	Yusri Yunus	36	3	Sendiri	Permanen	Mekanik	Menengah
39	Jufri Bai	45	5	Sendiri	Permanen	Petani	Menengah
40	Warni Datau	46	5	Sendiri	Permanen	Irt	Miskin
41	Rahman Putra Raden	29	3	Sendiri	Kayu	Petani	Menengah
42	Suriyono Boe	49	3	Sendiri	Kayu	Petani	Miskin
43	Burhan S Mauko	26	2	Orang Tua	Permanen	Petani	Menengah
44	Serly Ahmad Jaini	27	3	Sendiri	Permanen	Guru Honorer	Kaya
45	Idrak Rahman	43	5	Sendiri	Kayu	Buruh Bangunan	Miskin

(Sumber : Kantor Desa Dambalo, 2022)

Data yang digunakan pada penelitian ini adalah data masyarakat Desa Dambalo tahun 2022. Data tersebut meliputi Nama, Umur, Jumlah Tanggungan, Tempat Tinggal, Jenis Tempat Tinggal dan Pekerjaan serta Label Data yakni Kaya, Menengah, Miskin. Data yang digunakan sebagai *Data Uji* berjumlah 5 Data dari 45 *Data Training*.

4.2 Penjelasan Algoritma

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Metode *Bayes* merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining.

4.3 Perhitungan naïve bayes

Penulis menggunakan persamaan *Teorema Bayes* berikut:

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)}$$

Keterangan:

X : Kriteria suatu kasus berdasarkan masukan

C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas

(C_i|X) : Probabilitas kemunculan label kelas C_i dengan kriteria masukan X

P(X|C_i) : Probabilitas kriteria masukan X dengan label kelas

C_i P(C_i) : Probabilitas label kelas C_i

Alur kerja algoritma naïve bayes yaitu:

1. Menghitung jumlah kelas/label Langkah pertama yang harus dilakukan adalah menghitung jumlah kelas berdasarkan klasifikasi yang terbentuk (*prior probability*):

P(C_i)

$$P(C_{\text{Kaya}}) = 6/45 = 0,1333$$

$$P(C_{\text{Menengah}}) = 23/45 = 0,5111$$

$$P(C_{\text{Miskin}}) = 16/45 = 0,3555$$

2. Menghitung jumlah data per kelas/atribut yang ada.

P(X | Ci)

P(Tempat Tinggal|C_{Kaya, Menengah, Miskin})

$$P(\text{Sendiri}|C_{\text{Kaya}}) = 2/6 = 0,333$$

$$P(\text{Orang tua}|C_{\text{Kaya}}) = 3/6 = 0,5$$

$$P(\text{Numpang}|C_{\text{Kaya}}) = 1/6 = 0,166$$

$$P(\text{Sendiri}|C_{\text{Menengah}}) = 16/23 = 0,695$$

$$P(\text{Orang tua}|C_{\text{Menengah}}) = 7/23 = 0,304$$

$$P(\text{Numpang}|C_{\text{Menengah}}) = 0/23 = 0$$

$$P(\text{Sendiri}|C_{\text{Miskin}}) = 16/16 = 1$$

$$P(\text{Orang tua}|C_{\text{Miskin}}) = 0/16 = 0$$

$$P(\text{Numpang}|C_{\text{Miskin}}) = 0/16 = 0$$

P(Jenis Tempat Tinggal|C_{Kaya, Menengah, Miskin})

$$P(\text{Kayu}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Permanen}|C_{\text{Kaya}}) = 6/6 = 1$$

$$P(\text{Kayu}|C_{\text{Menengah}}) = 2/23 = 0,086$$

$$P(\text{Permanen}|C_{\text{Menengah}}) = 21/23 = 0,913$$

$$P(\text{Kayu}|C_{\text{Miskin}}) = 9/16 = 0,562$$

$$P(\text{Permanen}|C_{\text{Miskin}}) = 7/16 = 0,437$$

P(Pekerjaan|C_{Kaya, Menengah, Miskin})

$$P(\text{Petani}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Nelayan}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Honorar}|C_{\text{Kaya}}) = 1/6 = 0,166$$

$$P(\text{P.Swasta}|C_{\text{Kaya}}) = 1/6 = 0,166$$

$$P(\text{PNS}|C_{\text{Kaya}}) = 4/6 = 0,666$$

$$P(\text{IRT}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Buruh}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Perangkat Desa}|C_{\text{Kaya}}) = 0/6 = 0$$

$$P(\text{Petani} | C_{\text{Menengah}}) = 14/23 = 0,608$$

$$P(\text{Nelayan} | C_{\text{Menengah}}) = 2/23 = 0,086$$

$$P(\text{Honoror} | C_{\text{Menengah}}) = 0/23 = 0$$

$$P(\text{P.Swasta} | C_{\text{Menengah}}) = 1/23 = 0,043$$

$$P(\text{PNS} | C_{\text{Menengah}}) = 0/23 = 0$$

$$P(\text{IRT} | C_{\text{Menengah}}) = 2/23 = 0,086$$

$$P(\text{Buruh} | C_{\text{Menengah}}) = 3/23 = 0,130$$

$$P(\text{Perangkat Desa} | C_{\text{Menengah}}) = 1/23 = 0,043$$

$$P(\text{Petani} | C_{\text{Miskin}}) = 1/16 = 0,166$$

$$P(\text{Nelayan} | C_{\text{Miskin}}) = 3/16 = 0,187$$

$$P(\text{Honoror} | C_{\text{Miskin}}) = 0/16 = 0$$

$$P(\text{P.Swasta} | C_{\text{Miskin}}) = 0/16 = 0$$

$$P(\text{PNS} | C_{\text{Miskin}}) = 0/16 = 0$$

$$P(\text{IRT} | C_{\text{Miskin}}) = 5/16 = 0,312$$

$$P(\text{Buruh} | C_{\text{Miskin}}) = 1/16 = 0,166$$

$$P(\text{Perangkat Desa} | C_{\text{Miskin}}) = 0/16 = 0$$

P(Umur|C_{Kaya, Menengah, Miskin})

P(Tanggung|C_{Kaya, Menengah, Miskin})

Terdapat data numerik pada atribut Umur , maka temukan nilai mean dan standar deviasi dari masing-masing parameter yang menggambarkan data angka. Rumus yang digunakan untuk menghitung nilai rata – rata hitung (mean) dapat dilihat sebagai berikut :

Rumus menghitung mean

$$\mu = \sum_{i=1}^n x_i$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 \dots + x_n}{n}$$

$$\mu_{\text{umur} | C_{\text{Kaya}}} = 195/5 = 39$$

$$\mu_{\text{umur} | C_{\text{Menengah}}} = 758/19 = 39,894$$

$$\mu_{\text{umur}|C_{\text{Miskin}}} = 753/13 = 57,923$$

$$\mu_{\text{tanggungan}|C_{\text{kaya}}} = 5/2 = 2,5$$

$$\mu_{\text{tanggungan}|C_{\text{kaya}}} = 15/6 = 2,5$$

$$\mu_{\text{tanggungan}|C_{\text{kaya}}} = 21/6 = 3,5$$

Rumus jarak deviasi

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

$$\text{Umur}|C_{\text{Kaya}} = 12,72$$

$$\text{Umur}|C_{\text{Menengah}} = 3,79$$

$$\text{Umur}|C_{\text{Miskin}} = 4,400$$

$$\text{Tanggungan}|C_{\text{Kaya}} = 0,70$$

$$\text{Tanggungan}|C_{\text{Menengah}} = 1,87$$

$$\text{Tanggungan}|C_{\text{Miskin}} = 1,87$$

Contoh Data Uji :

Nama : IDRAK RAHMAN

Umur : 43

Tanggungan : 5

Tempat tinggal : Sendiri

Jenis Tempat Tinggal : Kayu

Pekerjaan : Buruh Bangunan

Status Ekonomi : ??

Penyelesaian :

Untuk mendapatkan hasil P(umur) dan P(Tanggungan) menggunakan rumus dibawah ini :

$$f(\omega) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\omega-\mu)^2}{2\sigma^2}}$$

$$P(\text{Umur}|C_{\text{kaya}}) = 0,140$$

$$P(\text{Umur}|C_{\text{menengah}}) = 1,249$$

$$P(\text{Umur}|C_{\text{miskin}}) = 1057,2$$

$$P(\text{Tanggung}|C_{\text{kaya}}) = 127338$$

$$P(\text{Tanggung}|C_{\text{Menengah}}) = 1,740$$

$$P(\text{Tanggung}|C_{\text{Miskin}}) = 0,555$$

1. $P(\text{Sendiri}|kaya) \times P(\text{Kayu}|kaya) \times (P(\text{Buruh}|kaya) \times P(\text{Umur}|kaya) \times P(\text{Tanggung}|kaya))$
2. $P(\text{Sendiri}|menengah) \times P(\text{Kayu}|menengah) \times (P(\text{Buruh}|menengah) \times P(\text{Umur}|menengah) \times P(\text{Tanggung}|menengah))$
3. $P(\text{Sendiri}|miskin) \times P(\text{Kayu}|miskin) \times (P(\text{Buruh}|miskin) \times P(\text{Umur}|miskin) \times P(\text{Tanggung}|miskin))$

Jawab

1. $0,33 \times 0 \times 0 \times 12,72 \times 0,70 = \mathbf{0 \text{ (Kaya)}}$
2. $0,695 \times 0,086 \times 0,130 \times 3,79 \times 1,87 = \mathbf{0,05 \text{ (Menengah)}}$
3. $1 \times 0,562 \times 0,166 \times 1057,2 \times 0,555 = \mathbf{54,73 \text{ (Miskin)}}$

Karena hasil perhitungan lebih besar **MISKIN**, maka data uji yang tersebut memiliki status ekonomi **MISKIN**.

4.4 Perhitungan Excel

4.4.1 Probilitas Tiap Label

Berdasarkan data masyarakat Desa Dambalo pada tabel 4.1 diketahui jumlah Dataset adalah sebanyak 45 data. Dimana 6 Data memiliki status ekonomi Kaya, 23 data memiliki status ekonomi menengah dan 16 Data berstatus Miskin seperti pada Tabel 4.2 berikut.

Tabel 4.2 Probilitas Tiap Label

P (Kaya/Menengah/Miskin)	Kaya	Menengah	Miskin
	6	23	16
Total Data	45		

4.4.2 Probilitas Kriteria Tempat Tinggal

Pada kriteria tempat tinggal dapat diketahui dari 45 data terdapat 2 Data berstatus Kaya memiliki Rumah Sendiri, 16 Data berstatus Menengah memiliki Rumah sendiri serta 16 Data bertatus Miskin memiliki Rumah Sendiri. Untuk yang tinggal bersama orang tua, 3 Data Kaya, 7 Data Menengah dan 0 Data Miskin. Untuk kriteria Numpang 1 Data Kaya seperti pada tabel 4.3 berikut.

Tabel 4.3 Probilitas Tempat Tinggal

P Tempat Tinggal	Jumlah Kejadian			Probilitas		
	Kaya	Menengah	Miskin	Kaya	Menengah	Miskin
Sendiri	2	16	16	0,3333333	0,695652174	1
Orang Tua	3	7	0	0,5	0,304347826	0
Numpang	1	0	0	0,1666667	0	0

4.4.3 Probilitas Kriteria Jenis Tempat Tinggal

Pada kriteria Jenis tempat tinggal dari 45 data terdapat 21 Data memiliki status ekonomi menengah bertempat tinggal permanen, 6 data berstatus menengah dan 7 data berstatus miskin. Selain itu, yang bertempat tinggal kayu, 3 data berstatus menengah dan 9 data berstatus miskin seperti pada tabel 4.4 berikut.

Tabel 4.4 Probilitas Jenis Tempat Tinggal

Jenis Tempat Tinggal	Jumlah Kejadian			Probilitas		
	Kaya	Menengah	Miskin	Kaya	Menengah	Miskin
Kayu	0	3	9	0	0,130	0,5625
Permanen	6	21	7	1	0,913043478	0,4375

4.4.4 Probilitas Kriteria Pekerjaan

Pada kriteria pekerjaan dari 45 data terdapat 14 data paling banyak memiliki status ekonomi menengah bekerja sebagai Petani serta 6 data berstatus

miskin bekerja sebagai petani. Selain itu, 4 data berstatus Kaya bekerja sebagai PNS seperti pada tabel 4.5 berikut ini.

Tabel 4.5 Probilitas Pekerjaan

P Pekerjaan	Jumlah Kejadian			Probilitas		
	Kaya	Menengah	Miskin	Kaya	Menengah	Miskin
Petani	0	14	6	0	0,608695652	0,375
Nelayan	0	2	3	0	0,086956522	0,1875
Honorar	1	0	0	0,166666667	0	0
Pegawai Swasta	1	1	0	0,166666667	0,043478261	0
PNS	4	0	0	0,666666667	0	0
IRT	0	2	5	0	0,086956522	0,3125
BURUH	0	3	1	0	0,130434783	0,0625
Perangkat Desa	0	1	0	0	0,043478261	0

4.4.5 Probilitas Kriteria Umur

Pada kriteria Umur karena memiliki nilai Numerik, maka nilai probilitas dapat diketahui dengan menentukan nilai mean dan jarak deviasi dari Kriteria umur seperti pada tabel 4.6 berikut.

Tabel 4.6 Probilitas Umur

P Umur	Kaya	Menengah	Miskin
	Mean	39	39,894
Standar Deviasi	12,7279	3,7964	4,4005

4.4.6 Probilitas Kriteria Tanggungan

Sama seperti Probilitas Kriteria Umur, probikitas tanggungan juga dapat diketahui degan menentukan nilai mean dan jarak deviasi dari kritria tanggungan seperti pada tabel 4.7 berikut.

Tabel 4.7 Probilitas Tanggungan

P Tanggungan	Kaya	Menengah	Miskin
Mean	2,5	2,5	3,5
Standar Deviasi	0,7071	1,8708	1,8708

4.4.7 Hasil Perhitungan Data Uji

Dalam menentukan hasil probabilitas dengan cara menghitung mana probabilitas yang paling besar. Probabilitas yang paling besar merupakan hasil akhir penentuan status ekonomi masyarakat. Hasil perhitungan probabilitas 5 Data Uji memberikan hasil Data 1,2,3,5 memiliki status ekonomi Miskin dan Data 4 memiliki status ekonomi Menengah seperti tabel 4.8 berikut ini.

Tabel 4.8 Hasil Perhitungan Data Uji

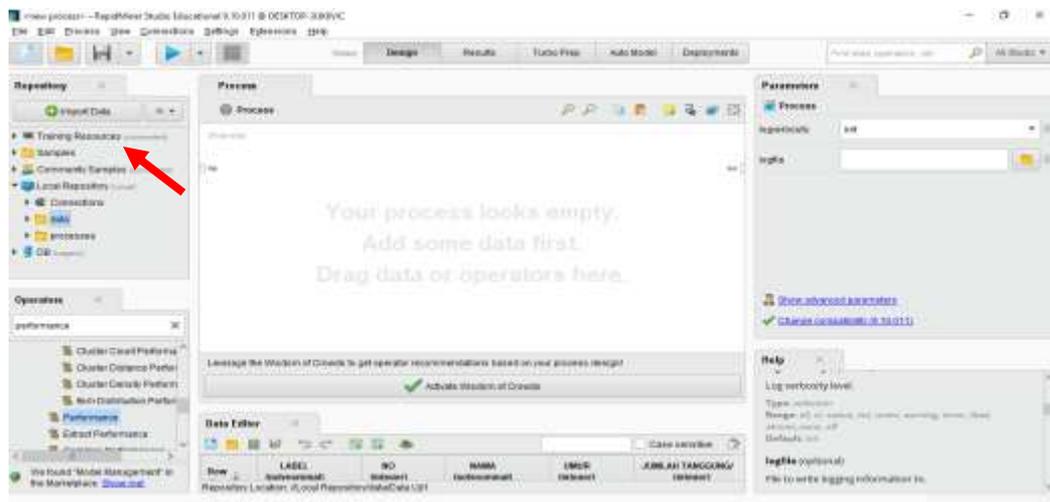
Data Uji	Hasil Perhitungan Probilitas			Status Ekonomi
	Kaya	Menengah	Miskin	
1	0	0	0,032	Miskin
2	0	0	0,01266	Miskin
3	0	0	8697,356	Miskin
4	0	2,61E+13	6,1563	Menengah
5	0	0,0995	1293,194	Miskin

BAB V

PEMBAHASAN PENELITIAN

5.1 Halaman Import Data

Halaman Import Data merupakan tampilan awal aplikasi Rapidminer. Halaman ini merupakan halaman yang digunakan untuk input data training dan data uji ke dalam aplikasi Rapidminer sebelum melakukan proses perhitungan algoritma Naïve Bayes.

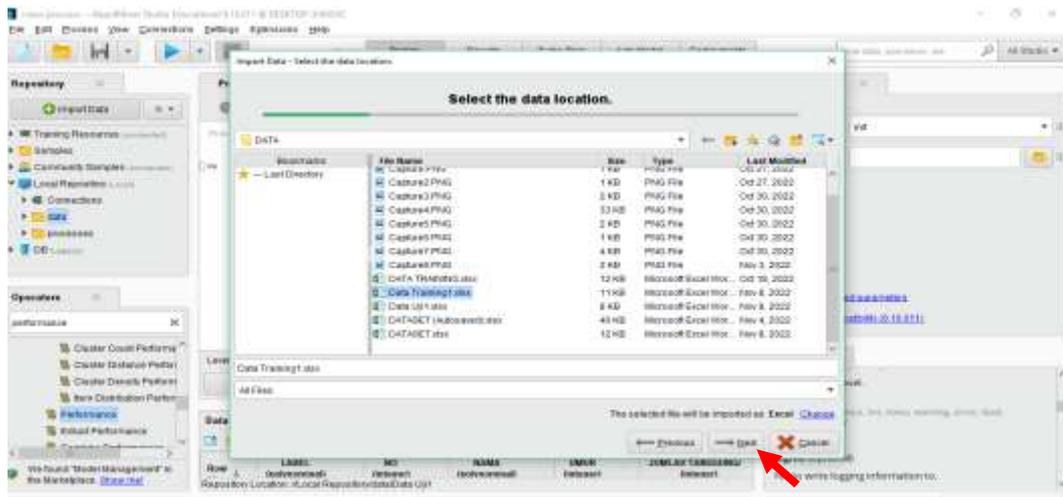


Gambar 5.1 Halaman Import Data

5.2 Tampilan Select Location Data

Setelah mengklik tombol Import Data pada halaman awal aplikasi Rapidminer, akan muncul halaman Select Location Data yang menampilkan keseluruhan data pada device komputer. Pilih Data Training untuk input Data Training.

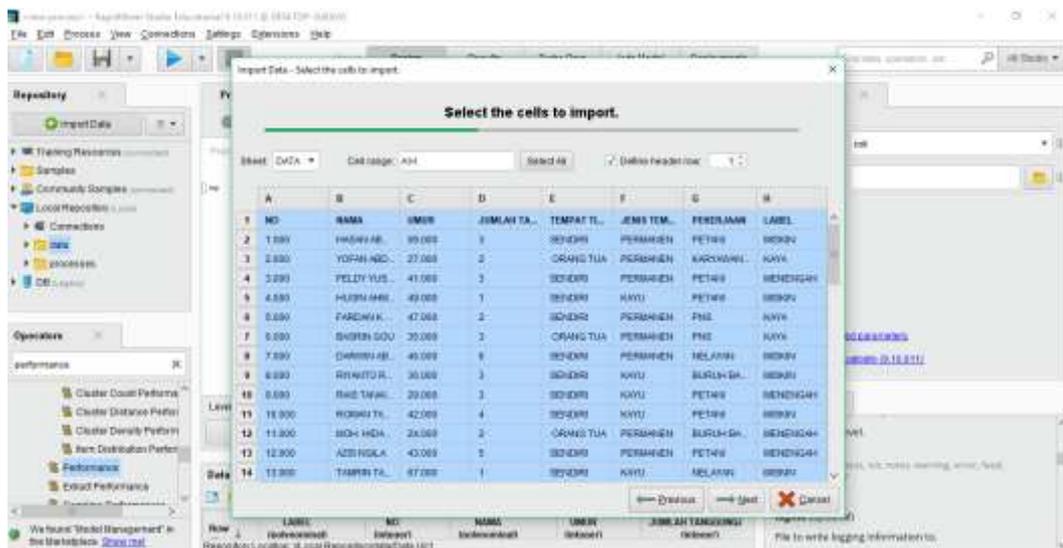
Setelah memilih Data Training kemudian pilih Tombol Next pada tampilan Selec Location Data seperti pada Gambar 5.2 berikut.



Gambar 5.2 Pilih Data Training dan Data Uji

5.3 Review Tampilan Data

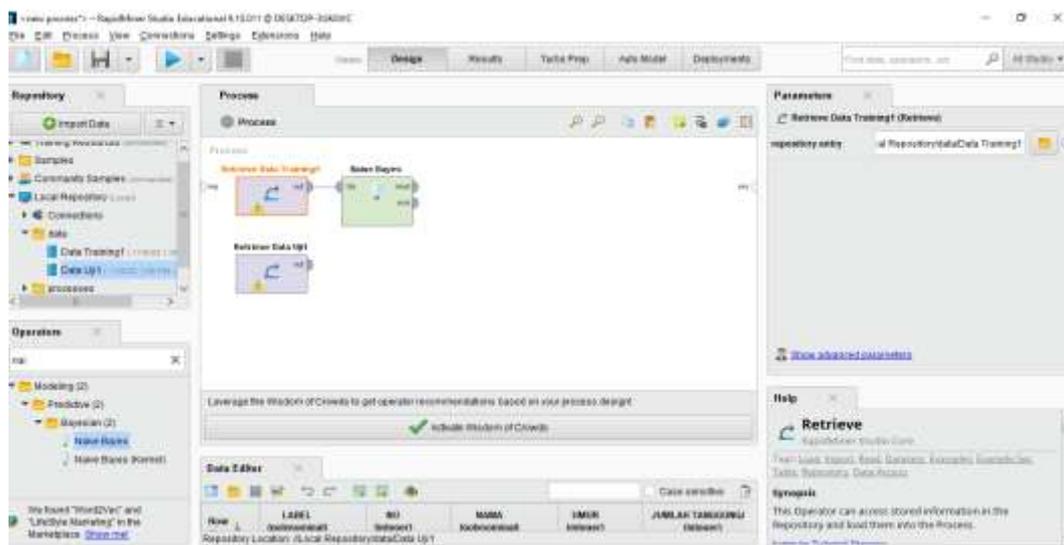
Setelah memilih Data Training akan muncul tampilan Halaman Review Data. Halaman ini digunakan untuk melihat kembali apakah data yang diinput sudah benar. Selain itu, di halaman ini juga digunakan untuk menghapus data yang tidak perlu dan menentukan label pada Data Training.



Gambar 5.3 Review Tampilan Data

5.4 Input Algoritma Naïve Bayes

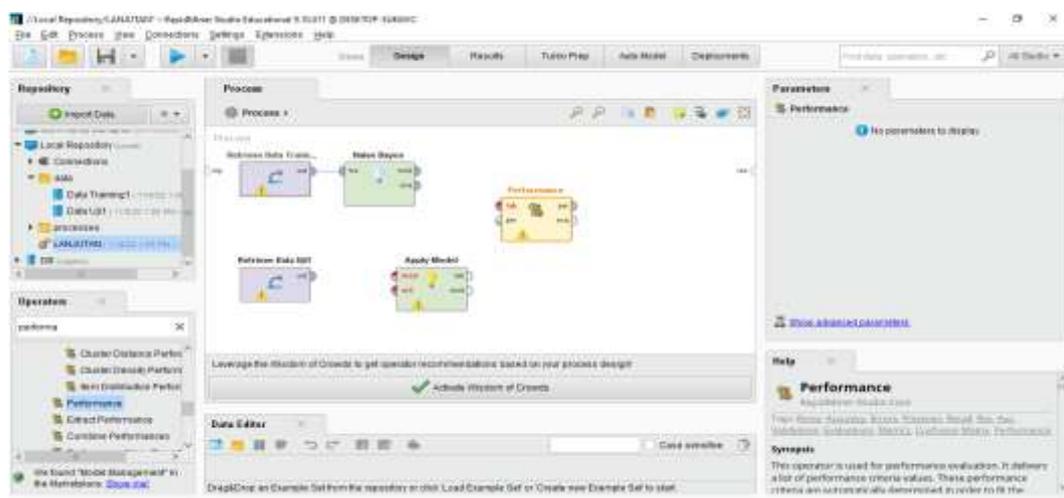
Setelah Data Training dan Data Uji dimasukkan dalam kolom proses, Masukkan Algoritma Naïve Bayes ke dalam kolom proses dengan cara mencari algoritma Naïve Bayes di bagian Operators.



Gambar 5.4 Input Algoritma

5.5 Input Apply Model dan Performance

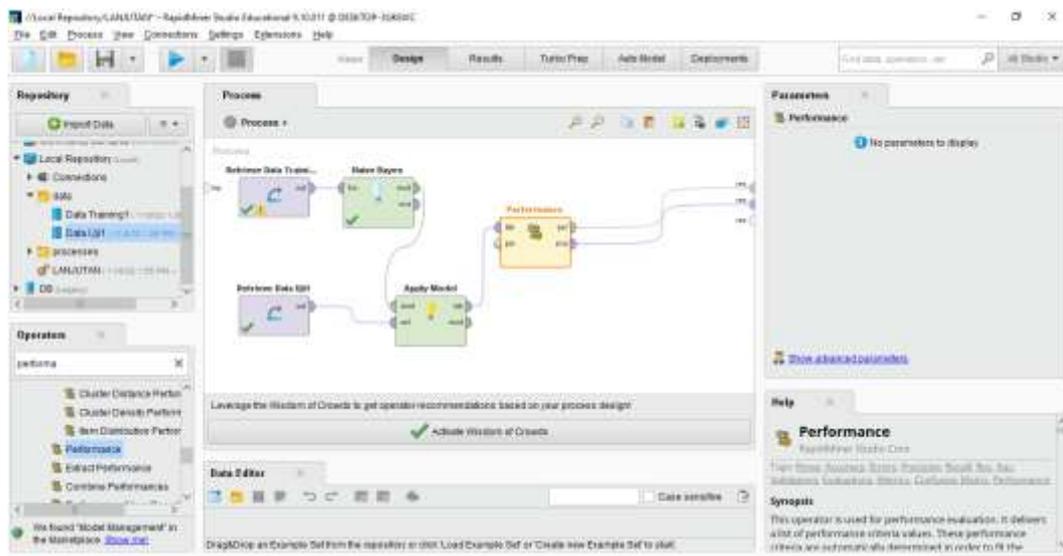
Setelah memasukkan Algoritma Naïve Bayes, selanjutnya ketik Apply Model dan Performane di bagian Operators kemudian masukan ke kolom proses seperti gambar berikut.



Gambar 5.5 Apply Model dan Performance

5.6 Hubungkan Semua Operators

Setelah *Apply Model* dan *Performance* dimasukkan ke dalam kolom proses, kemudian hubungkan Data Training ke *Naïve Bayes*, *Naïve Bayes* dihubungkan ke *Apply Model*, Data Uji dihubungkan Ke *Apply Model*, *Apply Model* ke *Performance* dan *Performance* dihubungkan ke Result seperti gambar berikut.



Gambar 5.6 Hubungkan Semua Operator

5.7 Proses Algoritma Naïve Bayes

Setelah dihubungkan Klik Icon Start untuk mulai proses perhitungan algoritma Naïve Bayes. Setelah itu muncul tampilan seperti gambar berikut. Performance Data Training berjumlah 45 data menggunakan rapidminer menghasilkan akurasi 100%.

	True MENENGAH	True KAYA	True MENENGAH	True precision
10	0	0	0	100.00%
0	0	0	0	100.00%
0	0	0	27	100.00%
	100.00%	100.00%	100.00%	

Gambar 5.7 Hasil Akurasi Data Training

5.8 Hasil Perhitungan Data Uji

Kemudian data uji berjumlah 5 data. Dari pengujian tersebut menghasilkan 4 Data memiliki Label Miskin, 1 Data memiliki Label Menengah seperti gambar berikut.

prediction	confidence	confidence	confidence	ID	NAME	SMSR	JMLAH DA
MISKIN	0.942	0.990	0.003	1	HARMA TANU	52	4
MISKIN	0.881	0.889	0.119	2	BARYATI RA	55	3
MISKIN	0.891	0.890	0.089	3	SATRA HEND	75	8
MENENGAH	0.120	0.023	0.042	4	HEDDAH AH	30	2
MISKIN	0.770	0.003	0.222	5	YANTI MUST	44	1

Gambar 5.8 Hasil Data Uji

BAB VI

PENUTUP

6.1 Kesimpulan

Setelah dilakukan penelitian di Kantor Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo Utara dan pembahasan yang telah diuraikan sebelumnya, maka dapat diperoleh kesimpulan bahwa:

1. Hasil kinerja algoritma Naïve Bayes untuk klasifikasi status ekonomi masyarakat Desa Dambalo Kecamatan Tomilito Kabupaten Gorontalo utara memperoleh akurasi 100%.
2. Hasil perhitungan 5 Data uji memperoleh hasil 4 data diantaranya memiliki status ekonomi **MISKIN** dan 1 data memiliki status ekonomi **MENENGAH**.

6.2 Saran

Setelah dilakukan penelitian dan pengujian dengan aplikasi rapidminer, terdapat beberapa saran yang harus diperhatikan untuk memperoleh tujuan yang diharapkan sebagai berikut:

1. Untuk penelitian selanjutnya perlu ditambahkan dataset dan variabel lain untuk mendapatkan hasil klasifikasi yang lebih baik lagi.
2. Penulis mengharapkan agar dilakukan penambahan data uji agar hasil klasifikasi lebih banyak.

DAFTAR PUSTAKA

- [1] R. F. K. Dewi, Obert, and R. Gusmana, "Implementasi Metode K-Nearest Neighbor (KNN) dalam Pengelompokan Status Ekonomi Warga," *J. Big Data Anal. Artif. Intell.*, vol. 4, no. 1, pp. 15–22, 2018.
- [2] N. Sastrawati, U. Islam, and N. Alauddin, "KONSUMTIVISME DAN STATUS SOSIAL," vol. 2, pp. 17–26, 2020.
- [3] R. T. Wulandari, "Pengertian Data Mining," *Data Min.*, vol. 7, no. 3, pp. 3–9, 2010.
- [4] W. Maharani, "Klasifikasi Data Menggunakan JST Backpropagation Momentum Dengan Adaptive Learning Rate," *Semin. Nas. Inform.*, vol. 1, no. semnasIF, pp. 25–31, 2009.
- [5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSl>.
- [6] Arifin, Z. (2019). Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql. *NJCA (Nusantara Journal of Computers and Its Applications)*, 4(1). <https://doi.org/10.36564/njca.v4i1.101>
- [7] Dewi, R. F. K., Obert, & Gusmana, R. (2018). Implementasi Metode K-Nearest Neighbor (KNN) dalam Pengelompokan Status Ekonomi Warga. *Journal of Big Data Analytic and Artificial Intelligence*, 4(1), 15–22.
- [8] Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., & Gata, W. (2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor). *Jurnal Techno Nusa Mandiri*, 16(1), 1–6. <https://doi.org/10.33480/techno.v16i1.25>
- [9] Lasena, Y. (2020). *CLUSTERING KOMODITI UNGGULAN MENGGUNAKAN ALGORITMA K-MEANS*. 2, 14–18.
- [10] Maharani, W. (2009). Klasifikasi Data Menggunakan JST

- Backpropagation Momentum Dengan Adaptive Learning Rate. *Seminar Nasional Informatika, 1*(semnasIF), 25–31.
- [11] Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTISI>
- [12] Nuris, M. (2015). White box testing pada sistem penilaian pembelajaran. *Skripsi Teknik Informatika*, 1–102.
- [13] Reza Noviansyah, M., Rismawan, T., & Marisa Midyanti, D. (2018). Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48–56. <http://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672>
- [14] Sastrawati, N., Islam, U., & Alauddin, N. (2020). *KONSUMTIVISME DAN STATUS SOSIAL*. 2, 17–26.
- [15] Wulandari, R. T. (2010). Pengertian Data Mining. *Data Mining*, 7(3), 3–9.

