

**PREDIKSI KONDISI PEMBAYARAN BIAYA PENDIDIKAN
SEKOLAH BERBASIS KLASIFIKASI EKONOMI SISWA**

THESIS

**Oleh :
DINI KRISTIANTI
NIM. 200605220015**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**PREDIKSI KONDISI PEMBAYARAN BIAYA PENDIDIKAN SEKOLAH
BERBASIS KLASIFIKASI EKONOMI SISWA**

THESIS

**Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam Memperoleh Gelar
Magister Ilmu Komputer (M.Kom)**

**Oleh :
DINI KRISTIANTI
NIM. 200605220015**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**PREDIKSI KONDISI PEMBAYARAN BIAYA PENDIDIKAN SEKOLAH
BERBASIS KLASIFIKASI EKONOMI SISWA**

THESIS

**Diajukan Kepada: Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan dalam Memperoleh
Gelar Magister Ilmu Komputer (M.Kom)**

**Oleh :
DINI KRISTIANTI
NIM. 200605220015**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS & TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

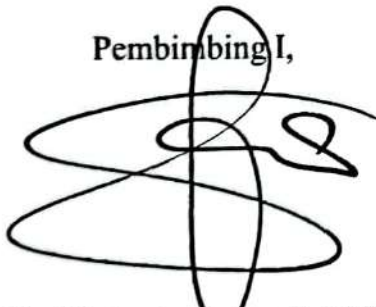
**PREDIKSI KONDISI PEMBAYARAN BIAYA PENDIDIKAN SEKOLAH
BERBASIS KLASIFIKASI EKONOMI SISWA**

THESIS

**Oleh :
DINI KRISTIANTI
NIM. 200605220015**

Telah diperiksa dan disetujui untuk diuji:
Tanggal:

Pembimbing I,



Dr. M. Amin Hariyadi, M.T
NIP. 19670118 200501 1 001

Pembimbing II,



Dr. Sahyo Crysdian
NIP. 19740424 200901 1 008

Mengetahui,
Dekan Fakultas Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Sahyo Crysdian
NIP. 19740424 200901 1 008

**PREDIKSI KONDISI PEMBAYARAN BIAYA PENDIDIKAN SEKOLAH
BERBASIS KLASIFIKASI EKONOMI SISWA**

THESIS

Oleh :
DINI KRISTIANTI
NIM. 200605220015

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal:

Susunan Dewan Penguji


Penguji Utama : Dr. Usman Pagalay, M.Si
196504142003121001

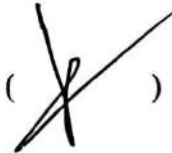
Ketua Penguji : Dr. Irwan Budi Santoso, M.Kom
197701032011011004

Sekretaris Penguji : Dr. M. Amin Hariyadi, M.T
196701182005011001

Anggota Penguji : Dr. Cahyo Crysdiان
197404242009011008

Tanda Tangan

()

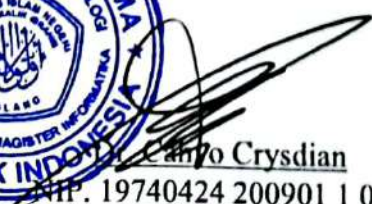
()

()

()

Mengetahui dan Mengesahkan
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Cahyo Crysdiان
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya, yang bertanda tangan di bawah ini:

Nama : Dini Kristianti

NIM : 200605220015

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Saya menyatakan dengan sejujurnya bahwa tesis yang saya buat ini benar-benar karya saya sendiri, bukan suatu pengadaan informasi, gubahan atau pertimbangan orang lain yang saya akui sebagai hasil karya atau perenungan saya sendiri, selain itu tesis ini saya buat dengan mengacu pada daftar pustaka.

Saya bersedia mengambil hukuman untuk hal-hal ini, apabila dikemudian hari terbukti atau dapat dibuktikan bahwa tesis ini adalah plagiat.

Malang,

Yang membuat pernyataan,



Dini Kristianti
NIM. 200605220015

MOTTO

Tidak ada ikhtiar yang sia-sia sama halnya dengan jangan takut akan kegagalan,
takutlah karena tidak mencoba.

LEMBAR PERSEMBAHAN

Thesis ini akan aku persembahkan untuk kedua orang tuaku, Bapak (Alm. Supeno) dan Ibu (Rukiyani Dayanti), suami (Akhmad Yusuf) juga adikku tercinta (Mahfirani Ilusia Insiroh) serta kedua anakku tersayang.

Terima kasih atas dukungan dan doa yang tak pernah ada hentinya.

Dan tak lupa ucapan terima kasih kepada teman-teman seperjuanganku yang telah melewatkan waktu bersama saat suka, duka selama kuliah.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya sehingga saya dapat menyelesaikan thesis ini dan menyelesaikan studinya di Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. .

Selain itu, penulis ingin menyampaikan penghargaan, doa dan harapannya, kepada semua pihak yang telah membantu penyelesaian thesis. Penulis mengucapkan terimakasih kepada:

1. Selaku pembimbing thesis, Bapak Dr. M. Amin Hariyadi dan Dr. Cahyo Crysdian yang telah memberikan bimbingan dan pengalaman yang sangat berharga.
2. Saya mengucapkan terima kasih kepada seluruh civitas akademika yang tergabung dalam Program Studi Magister Informatika, khususnya kepada seluruh dosen.
3. Ibu dan suami tercinta yang selalu mendoakan dan menyemangati penulis dalam menempuh pendidikan,
4. Semua pihak yang telah memberikan kontribusi, moral dan materiil, untuk penyelesaian thesis ini.

Penulis berharap untuk thesis ini dapat memberikan manfaat bagi para pembaca, khususnya penulis sendiri, meskipun masih terdapat kekurangan dalam proses penyusunannya. Amin ya, Rabbal Alamin

Wassalamu'alaikum, Wr. Wb.

Malang,
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN.....	v
MOTTO	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xvi
<i>Abstrak</i>	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar belakang	1
1.1. Pernyataan Masalah	7
1.2. Tujuan Penelitian.....	7
1.3. Manfaat Penelitian.....	7
1.4. Batasan Masalah	7
1.5. Sistematika Penulisan.....	7
BAB II STUDI PUSTAKA	9
2.1. Sistem Prediksi Biaya Pendidikan Sekolah	9
2.2. Kerangka Teori	12
BAB III DESAIN PENELITIAN.....	16
3.1. Desain Penelitian	16
3.1.1. Perencanaan	17
3.1.2. Tindakan / Pelaksanaan	27
3.1.3. Pengamatan	27
BAB IV METODE <i>NAIVE BAYES</i>	29
4.1. Desain.....	29
4.2. Uji Coba	34
4.2.1. Percobaan Pertama Data Uji 80:20	34

4.2.2.	Percobaan Kedua Data Uji 75:25	52
4.2.3.	Percobaan Ketiga Data Uji 70:30	57
4.2.4.	Percobaan Keempat Data Uji 60:40	63
4.2.5.	Percobaan Kelima Data Uji 50:50	68
4.3.	Kesimpulan.....	74
BAB V METODE <i>DECISION TREE</i>		77
5.1.	Desain.....	77
5.2.	Uji Coba	81
5.2.1.	Percobaan Pertama Data Uji 80:20	82
5.2.2.	Percobaan Kedua Data Uji 75:25	98
5.2.3.	Percobaan Ketiga Data Uji 70:30	105
5.2.4.	Percobaan Keempat Data Uji 60:40	113
5.2.5.	Percobaan Kelima Data Uji 50:50	120
5.3.	Kesimpulan.....	128
BAB VI PEMBAHASAN		131
6.1.	Pembahasan Komparasi Algoritma	131
BAB VII KESIMPULAN DAN SARAN		139
7.1.	Kesimpulan.....	139
7.2.	Saran	140
DAFTAR PUSTAKA		141
LAMPIRAN-LAMPIRAN		143

DAFTAR TABEL

Tabel 2. 1 Daftar Jurnal	13
Tabel 3. 1 Data siswa	17
Tabel 3. 2 Tabel data siswa	18
Tabel 3. 3 Data orang tua	19
Tabel 3. 4 Tabel Atribut data orang tua	20
Tabel 3. 5 Data siswa sudah di normalisasi	20
Tabel 3. 6 Data orang tua sudah di normalisasi	21
Tabel 3. 7 Data sudah di normalisasi	22
Tabel 3. 8 Data selection	23
Tabel 3. 9 Kategori kriteria gaji	24
Tabel 3. 10 Kategori tanggungan orang tua	25
Tabel 3. 11 Kategori pekerjaan orang tua	25
Tabel 3. 12 Data kondisi siswa	26
Tabel 4. 1 Pekerjaan Orang Tua	31
Tabel 4. 2 Penghasilan orang tua per bulan	32
Tabel 4. 3 Jumlah Tanggungan	32
Tabel 4. 4 Jumlah dataset uji coba 80:20 skenario pertama	35
Tabel 4. 5 Probabilitas actual class uji coba 80:20 skenario pertama	35
Tabel 4. 6 Pekerjaan orang tua uji coba 80:20 skenario pertama	35
Tabel 4. 7 Penghasilan orang tua per bulan uji coba 80:20 skenario pertama	36
Tabel 4. 8 Jumlah Tanggungan uji coba 80:20 skenario pertama	36
Tabel 4. 9 <i>Confusion matrix</i>	38
Tabel 4. 10 Hasil <i>confusion matrix</i> uji coba 80:20 skenario pertama	39
Tabel 4. 11 Hasil uji coba 80:20 skenario pertama	40
Tabel 4. 12 Jumlah dataset uji coba 80:20 skenario kedua	40
Tabel 4. 13 Probabilitas actual class uji coba 80:20 skenario kedua	40
Tabel 4. 14 Pekerjaan orang tua uji coba 80:20 skenario kedua	41
Tabel 4. 15 Penghasilan orang tua per bulan uji coba 80:20 skenario kedua	41
Tabel 4. 16 Hasil <i>confusion matrix</i> percobaan 80:20 skenario kedua	43
Tabel 4. 17 Hasil uji coba 80:20 skenario kedua	44
Tabel 4. 18 Jumlah dataset uji coba 80:20 skenario ketiga	44
Tabel 4. 19 Probabilitas actual class uji coba 80:20 skenario ketiga	44
Tabel 4. 20 Pekerjaan orang tua uji coba 80:20 skenario ketiga	45
Tabel 4. 21 Jumlah Tanggungan uji coba 80:20 skenario ketiga	45
Tabel 4. 22 Hasil <i>confusion matrix</i> uji coba 80:20 skenario ketiga	47
Tabel 4. 23 Hasil uji coba 80:20 skenario ketiga	47
Tabel 4. 24 Jumlah dataset uji coba 80:20 skenario keempat	48
Tabel 4. 25 Probabilitas actual class uji coba 80:20 skenario keempat	48
Tabel 4. 26 Penghasilan orang tua per bulan uji coba 80:20 skenario keempat	49
Tabel 4. 27 Jumlah Tanggungan uji coba 80:20 skenario keempat	49
Tabel 4. 28 Hasil <i>confusion matrix</i> uji coba 80:20 skenario keempat	51
Tabel 4. 29 Hasil uji coba 80:20 skenario keempat	51
Tabel 4. 30 Hasil <i>confusion matrix</i> uji coba 75:25 skenario pertama	52

Tabel 4. 31 Hasil uji coba 75:25 skenario pertama.....	53
Tabel 4. 32 Hasil <i>confusion matrix</i> uji coba 75:25 skenario kedua	54
Tabel 4. 33 Hasil uji coba 75:25 skenario kedua	54
Tabel 4. 34 Hasil <i>confusion matrix</i> uji coba 75:25 skenario ketiga	55
Tabel 4. 35 Hasil uji coba 75:25 skenario ketiga.....	56
Tabel 4. 36 Hasil <i>confusion matrix</i> uji coba 75:25 skenario keempat	56
Tabel 4. 37 Hasil uji coba 75:25 skenario keempat	57
Tabel 4. 38 Hasil <i>confusion matrix</i> uji coba 70:30 skenario pertama.....	58
Tabel 4. 39 Hasil uji coba 70:30 skenario pertama.....	59
Tabel 4. 40 Hasil <i>confusion matrix</i> uji coba 70:30 skenario kedua	59
Tabel 4. 41 Hasil uji coba 70:30 skenario kedua	60
Tabel 4. 42 Hasil <i>confusion matrix</i> uji coba 70:30 skenario ketiga	61
Tabel 4. 43 Hasil uji coba 70:30 skenario ketiga.....	61
Tabel 4. 44 Hasil <i>confusion matrix</i> uji coba 70:30 skenario keempat	62
Tabel 4. 45 Hasil uji coba 70:30 skenario keempat	63
Tabel 4. 46 Hasil <i>confusion matrix</i> uji coba 60:40 skenario pertama.....	63
Tabel 4. 47 Hasil uji coba 60:40 skenario pertama.....	64
Tabel 4. 48 Hasil <i>confusion matrix</i> uji coba 60:40 skenario kedua	65
Tabel 4. 49 Hasil uji coba 60:40 skenario kedua	66
Tabel 4. 50 Hasil <i>confusion matrix</i> uji coba 60:40 skenario ketiga.....	66
Tabel 4. 51 Hasil uji coba 60:40 skenario ketiga.....	67
Tabel 4. 52 Hasil <i>confusion matrix</i> uji coba 60:40 skenario keempat	68
Tabel 4. 53 Hasil uji coba 60:40 skenario keempat	68
Tabel 4. 54 Hasil <i>confusion matrix</i> uji coba 50:50 skenario pertama.....	69
Tabel 4. 55 Hasil uji coba 50:50 skenario pertama.....	70
Tabel 4. 56 Hasil <i>confusion matrix</i> uji coba 50:50 skenario kedua	70
Tabel 4. 57 Hasil uji coba 50:50 skenario kedua	71
Tabel 4. 58 Hasil <i>confusion matrix</i> uji coba 50:50 skenario ketiga	72
Tabel 4. 59 Hasil uji coba 50:50 skenario ketiga.....	73
Tabel 4. 60 Hasil <i>confusion matrix</i> uji coba 50:50 skenario keempat	73
Tabel 4. 61 Hasil uji coba 50:50 skenario keempat	74
Tabel 5. 1 Klasifikasi berdasarkan pekerjaan orang tua	80
Tabel 5. 2 Klasifikasi berdasarkan penghasilan orang tua per bulan.....	80
Tabel 5. 3 Klasifikasi berdasarkan jumlah tanggungan	80
Tabel 5. 4 Node awal	81
Tabel 5. 5 Jumlah dataset uji coba 80:20 skenario pertama.....	82
Tabel 5. 6 Klasifikasi pekerjaan orang tua uji coba 80:20 skenario pertama.....	83
Tabel 5. 7 Klasifikasi penghasilan orang tua uji coba 80:20 skenario pertama	83
Tabel 5. 8 Klasifikasi jumlah tanggungan uji coba 80:20 skenario pertama	83
Tabel 5. 9 Node awal uji coba 80:20 skenario pertama	84
Tabel 5. 10 <i>Confusion matrix</i>	85
Tabel 5. 11 Hasil <i>confusion matrix</i> uji coba 80:20 skenario pertama.....	85
Tabel 5. 12 Hasil uji coba 80:20 skenario pertama.....	86
Tabel 5. 13 Jumlah dataset uji coba 80:20 skenario kedua	87

Tabel 5. 14	Klasifikasi pekerjaan orang tua uji coba 80:20 skenario kedua	87
Tabel 5. 15	Klasifikasi penghasilan orang tua uji coba 80:20 skenario kedua	88
Tabel 5. 16	Node awal uji coba 80:20 skenario kedua	88
Tabel 5. 17	Hasil <i>confusion matrix</i> uji coba 80:20 skenario kedua	89
Tabel 5. 18	Hasil uji coba 80:20 skenario kedua	90
Tabel 5. 19	Jumlah dataset uji coba 80:20 skenario ketiga.....	91
Tabel 5. 20	Klasifikasi pekerjaan orang tua uji coba 80:20 skenario ketiga.....	91
Tabel 5. 21	Klasifikasi jumlah tanggungan uji coba 80:20 skenario ketiga	91
Tabel 5. 22	Node awal uji coba 80:20 skenario ketiga	92
Tabel 5. 23	Hasil <i>confusion matrix</i> uji coba 80:20 skenario ketiga	93
Tabel 5. 24	Hasil Uji coba 80:20 skenario ketiga	93
Tabel 5. 25	Jumlah dataset uji coba 80:20 skenario keempat	94
Tabel 5. 26	Klasifikasi penghasilan orang tua uji coba 80:20 skenario keempat	95
Tabel 5. 27	Klasifikasi jumlah tanggungan uji coba 80:20 skenario keempat.....	95
Tabel 5. 28	Node awal uji coba 80:20 skenario keempat	95
Tabel 5. 29	Hasil <i>confusion matrix</i> uji coba 80:20 skenario keempat	96
Tabel 5. 30	Hasil uji coba 80:20 skenario keempat	97
Tabel 5. 31	Hasil <i>confusion matrix</i> uji coba 75:25 skenario pertama.....	98
Tabel 5. 32	Hasil uji coba 75:25 skenario pertama.....	99
Tabel 5. 33	Hasil <i>confusion matrix</i> uji coba 75:25 skenario kedua	100
Tabel 5. 34	Hasil uji coba 75:25 skenario kedua	101
Tabel 5. 35	Hasil <i>confusion matrix</i> uji coba 75:25 skenario ketiga	102
Tabel 5. 36	Hasil uji coba 75:25 skenario ketiga.....	102
Tabel 5. 37	Hasil <i>confusion matrix</i> uji coba 75:25 skenario keempat	104
Tabel 5. 38	Hasil uji coba 75:25 skenario keempat	104
Tabel 5. 39	Hasil <i>confusion matrix</i> uji coba 70:30 skenario pertama.....	106
Tabel 5. 40	Hasil uji coba 70:30 skenario pertama.....	106
Tabel 5. 41	Hasil <i>confusion matrix</i> uji coba 70:30 skenario kedua	107
Tabel 5. 42	Hasil uji coba 70:30 skenario kedua	108
Tabel 5. 43	Hasil <i>confusion matrix</i> uji coba 70:30 skenario ketiga	109
Tabel 5. 44	Hasil uji coba 70:30 skenario ketiga.....	110
Tabel 5. 45	Hasil <i>confusion matrix</i> uji coba 70:30 skenario keempat	111
Tabel 5. 46	Hasil uji coba 70:30 skenario keempat	112
Tabel 5. 47	Hasil <i>confusion matrix</i> uji coba 60:40 skenario pertama.....	113
Tabel 5. 48	Hasil uji coba 60:40 skenario pertama.....	114
Tabel 5. 49	Hasil <i>confusion matrix</i> uji coba 60:40 skenario kedua	115
Tabel 5. 50	Hasil uji coba 60:40 skenario kedua	116
Tabel 5. 51	Hasil <i>confusion matrix</i> uji coba 60:40 skenario ketiga	117
Tabel 5. 52	Hasil uji coba 60:40 skenario ketiga.....	118
Tabel 5. 53	Hasil <i>confusion matrix</i> uji coba 60:40 skenario keempat	119
Tabel 5. 54	Hasil uji coba 60:40 skenario keempat	120
Tabel 5. 55	Hasil <i>confusion matrix</i> uji coba 50:50 skenario pertama.....	121
Tabel 5. 56	Hasil uji coba 50:50 skenario pertama.....	122
Tabel 5. 57	Hasil <i>confusion matrix</i> uji coba 50:50 skenario kedua	123

Tabel 5. 58 Hasil uji coba 50:50 skenario kedua	123
Tabel 5. 59 Hasil <i>confusion matrix</i> uji coba 50:50 skenario ketiga	125
Tabel 5. 60 Hasil uji coba 50:50 skenario ketiga	125
Tabel 5. 61 Hasil <i>confusion matrix</i> uji coba 50:50 skenario keempat	126
Tabel 5. 62 Hasil uji coba 50:50 skenario keempat	127
Tabel 6. 1 Hasil rasio pekerjaan orang tua.....	132
Tabel 6. 2 Hasil rasio penghasilan orang tua	132
Tabel 6. 3 Hasil rasio jumlah tanggungan orang tua.....	133

DAFTAR GAMBAR

Gambar 2. 1 Kerangka teori.....	12
Gambar 3. 1 Desain Penelitian.....	16
Gambar 3. 2 Desain Sistem.....	27
Gambar 3. 3 Konsep model	27
Gambar 4. 1 Desain penelitian naïve bayes	29
Gambar 5. 1 Desain penelitian decision tree.....	77
Gambar 5. 2 Node akar percobaan 80:20 skenario pertama	86
Gambar 5. 3 Node akar percobaan 80:20 skenario kedua.....	90
Gambar 5. 4 Node akar percobaan 80:20 skenario ketiga	94
Gambar 5. 5 Node akar percobaan 80:20 skenario keempat.....	97
Gambar 5. 6 Node akar percobaan 75:25 skenario pertama	99
Gambar 5. 7 Node akar percobaan 75:25 skenario kedua.....	101
Gambar 5. 8 Node akar percobaan 75:25 skenario ketiga	103
Gambar 5. 9 Node akar percobaan 75:25 skenario keempat.....	105
Gambar 5. 10 Node akar percobaan 70:30 skenario pertama	107
Gambar 5. 11 Node akar percobaan 70:30 skenario kedua.....	109
Gambar 5. 12 Node akar percobaan 70:30 skenario ketiga	111
Gambar 5. 13 Node akar percobaan 70:30 skenario keempat.....	113
Gambar 5. 14 Node akar percobaan 60:40 skenario pertama	115
Gambar 5. 15 Node akar percobaan 60:40 skenario kedua.....	116
Gambar 5. 16 Node akar percobaan 60:40 skenario ketiga	118
Gambar 5. 17 Node akar percobaan 60:40 skenario keempat.....	120
Gambar 5. 18 Node akar percobaan 50:50 skenario pertama	122
Gambar 5. 19 Node akar percobaan 50:50 skenario kedua.....	124
Gambar 5. 20 Node akar percobaan 50:50 skenario ketiga	126
Gambar 5. 21 Node akar percobaan 50:50 skenario keempat.....	128

Abstrak

Biaya pendidikan yang dibayarkan siswa merupakan salah satu sumber dana tetap dalam meningkatkan kualitas pendidikan bagi sekolah swasta. Dalam melakukan transaksi pembayaran biaya pendidikan, apabila para siswa menunggak dalam membayar biaya pendidikan, maka hal tersebut menjadi masalah karena dapat menghambat proses kegiatan belajar mengajar di sekolah. Oleh karena itu, perlu adanya evaluasi yang dilakukan dalam pembayaran biaya pendidikan. Untuk mengatasi permasalahan tersebut perlu dilakukan pengelompokan kondisi ekonomi siswa dengan menggunakan data mining. Penelitian tersebut dilakukan dengan menggunakan beberapa variabel seperti jumlah pendapatan, tanggungan keluarga, pekerjaan orang tua, dengan menggunakan data kondisi siswa. Data yang ada dianalisis diimplementasikan kedalam bentuk pemrograman menggunakan bahasa pemrograman python dengan algoritma Naïve Bayes dan Decision Tree. Naïve Bayes dan Decision Tree merupakan salah satu metode pada Probabilistic Reasoning. Hasil akurasi yang diperoleh algoritma Naïve Bayes dan Decision Tree adalah 71%.

Kata kunci: *Data Mining, Klasifikasi, Algoritma Naïve Bayes, Algoritma Decision Tree, Perhitungan Python, Biaya pendidikan, Iuran Pendidikan Sekolah*

Abstract

One way that private schools can raise fixed funds to improve education quality is through student tuition fees. If students are behind on their education costs, it's a problem because it can slow down teaching and learning in schools. When transactions are made to pay for education, this can cause problems. As a result, a payment evaluation for educational expenses must be carried out. Using data mining, it is necessary to group students' economic circumstances in order to solve this issue. The examination was done utilizing a few factors, for example, pay, family wards, parental work, utilizing understudy condition information. Using the python programming language and the Naive Bayes and Decision Tree algorithms, the analyzed existing data is programmed. One approach to probabilistic reasoning is the use of novice bayes and decision trees. The Naive Bayes and Decision Tree algorithms achieve an accuracy of 71%.

Keywords: *Data Mining, Classification, the Decision Tree Algorithm, the Python Calculation, the Naive Bayes Algorithm, and Education Costs*

نبذة مختصرة

تعتبر الرسوم الدراسية التي يدفعها الطلاب من المصادر الدائمة للتمويل في تحسين جودة التعليم للمدارس الخاصة. عند إجراء معاملات الدفع لرسوم التعليم ، إذا تأخر الطلاب في دفع الرسوم الدراسية ، فإن هذا يصبح مشكلة لأنه يمكن أن يعيق عملية أنشطة التدريس والتعلم في المدرسة. لذلك ، هناك حاجة لإجراء تقييم في دفع الرسوم الدراسية. للتغلب على هذه المشاكل ، من الضروري تجميع الظروف الاقتصادية للطلاب باستخدام التنقيب عن البيانات. تم إجراء البحث باستخدام عدة متغيرات مثل مقدار الدخل ، والمُعاليين من الأسرة ، وعمل الوالدين ، باستخدام بيانات مع الثعابين عن ظروف الطالب. يتم تحليل البيانات الحالية وتنفيذها في شكل برمجة باستخدام لغة برمجة هي إحدى طرق التفكير الاحتمالي. بلغت نتائج الدقة التي حصلت عليها كلا الطريقتين. نايف بايز و خوارزميات وشجرة القرار 71٪ خوارزمية باينز ساذجة، خوارزمية شجرة القرار ، حسابات الكلمات الرئيسية: استخراج البيانات ، التصنيف ، خوارزمية ، تكاليف التعليم ، الرسوم المدرسية الثعابين

BAB I

PENDAHULUAN

1.1.Latar belakang

Bidang pendidikan salah satu contoh peranan penting untuk memajukan bangsa. Dunia pendidikan merupakan tempat dimana putra putri bangsa mendapatkan ilmu, wawasan dan pendidikan karakter. Pendidikan formal dimulai dari tingkat SD, SMP dan SMA/SMK. Untuk menunjang proses belajar mengajar di sekolah, perlu adanya sebuah biaya pendidikan.

Salah satu kewajiban untuk mendapatkan pendidikan di sekolah adalah melakukan pembayaran biaya pendidikan. Tidak dapat dipungkiri, biaya pendidikan merupakan salah satu faktor penting yang digunakan untuk mengalokasikan biaya pendidikan pembangunan sekolah, untuk guru, karyawan, dan lain-lain. Biaya pendidikan ini umumnya diterapkan oleh sekolah swasta yang dibebankan kepada siswanya. Biaya pendidikan ini dibayarkan setiap bulan.

Dalam hal ini masalah yang muncul terkait dalam pembayaran biaya pendidikan yaitu apabila para siswa menunggak dan tidak membayar biaya pendidikan. Hal tersebut menjadi masalah karena pembayaran biaya pendidikan merupakan salah satu sumber dana tetap dalam meningkatkan kualitas pendidikan sekolah. Hal itu disebabkan oleh beberapa faktor seperti : jumlah pendapatan, tanggungan keluarga, pekerjaan orang tua. Hal tersebut menjadi masalah yang besar karena banyaknya siswa yang terlambat dalam membayar biaya pendidikan akan menurunkan pemasukan dana tetap sekolah. Maka perlu adanya solusi berupa sistem klasifikasi kondisi ekonomi siswa, yang dimana dapat diketahui faktor apa yang mempengaruhi tunggakan siswa dalam melakukan pembayaran biaya pendidikan.

Pembiayaan pendidikan pada dasarnya lahir dari pengertian yaitu “saya percaya” atau “saya menaruh kepercayaan”. Perkataan pembiayaan pendidikan yang artinya kepercayaan yang berarti walimurid menaruh kepercayaan kepada seseorang untuk melaksanakan amanah yang diberikan oleh walimurid selaku

shahibul maal. Dana tersebut harus digunakan dengan benar, adil, dan harus disertai dengan ikatan dan syarat-syarat yang jelas, sebagaimana firman Allah SWT dalam: Q.S. An-Nisa“ (4) ayat 29:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُم بَيْنَكُم بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ ۗ وَلَا تَقْتُلُوا أَنْفُسَكُمْ ۗ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا ۙ ٢٩

Artinya: “Hai orang-orang yang beriman, janganlah kamu saling memakan harta sesamamu dengan jalan yang batil, kecuali dengan jalan perniagaan yang berlaku dengan suka sama suka di antara kamu. Dan janganlah kamu membunuh dirimu; sesungguhnya Allah adalah Maha Penyayang kepadamu”. (Q.S. An-Nisa: 29)

Pada ayat tersebut menurut tafsir Al-Jalalain “(Hai orang-orang yang beriman! Janganlah kamu makan harta sesamamu dengan jalan yang batil) artinya jalan yang haram menurut agama seperti riba dan gasab/merampas (kecuali dengan jalan) atau terjadi (secara perniagaan) menurut suatu qiraat dengan baris di atas sedangkan maksudnya ialah hendaklah harta tersebut harta perniagaan yang berlaku (dengan suka sama suka di antara kamu) berdasar kerelaan hati masing-masing, maka bolehlah kamu memakannya. (Dan janganlah kamu membunuh dirimu) artinya dengan melakukan hal-hal yang menyebabkan kecelakaannya bagaimana pun juga cara dan gejalanya baik di dunia dan di akhirat. (Sesungguhnya Allah Maha Penyayang kepadamu) sehingga dilarang-Nya kamu berbuat demikian”. Ayat tersebut menyuruh kita untuk menginfakkan dan membelanjakan harta kita semata-mata untuk mencari keridhoan Allah Swt. Salah satunya pada bidang pendidikan dengan biaya pendidikan harus direncanakan dan mencari keridhoan Allah SWT.

Menurut surat al-Mujadilah ayat 12–13, Allah SWT secara implisit telah menyebutkan masalah pembiayaan pendidikan:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا نَاجَيْتُمُ الرَّسُولَ فَقَدِّمُوا بَيْنَ يَدَيْ نَجْوٰكُمُ صَدَقَةً ۚ ذٰلِكَ خَيْرٌ لَّكُمْ وَأَطْهَرٌ ۚ فَإِنْ لَّمْ تَجِدُوا فَإِنَّ اللَّهَ غَفُورٌ رَّحِيمٌ (١٢)

ءَأَشْفَقْتُمْ أَنْ تُقَدِّمُوا بَيْنَ يَدَيْ نَجْوٰكُمُ صَدَقَاتٍ ۚ فَإِذْ لَمْ تَفْعَلُوا وَتَابَ اللَّهُ عَلَيْكُمْ فَأَقْبِمُوا الصَّلٰوةَ وَآتُوا الزَّكٰوةَ وَاطِيعُوا اللَّهَ وَرَسُوْلَهُ ۗ وَاللَّهُ خَبِيرٌ ۙ بِمَا تَعْمَلُونَ (١٣)

Artinya adalah: "Wahai orang-orang yang beriman, sebelum melakukan percakapan khusus dengan Rasulullah, Anda harus memberi sedekah kepada orang miskin. Itu lebih bersih dan lebih baik untuk Anda. Namun, sesungguhnya Allah Maha Pengampun lagi Maha Penyayang jika kamu tidak menerima (yang akan disedekahkan). Apakah Anda khawatir menjadi miskin karena Anda memberi uang sebelum berbicara dengan Rasul? Namun, jika Anda gagal lakukan, dan Tuhan menyediakan Anda pengampunan, maka berdoa, berzakat, dan ikutilah Allah dan Rasul-Nya! Dan Allah mengetahui segala sesuatu tentang apa yang kamu kerjakan".

Ayat tersebut menurut tafsir Al-Jalalain "Hai orang-orang yang beriman, jika hendak mengadakan pembicaraan khusus dengan rasul, hendaknya mengeluarkan sebelum pembicaraan (sedekah) kalian sebelum diadakan pembicaraan khusus. Itu lebih baik bagi kalian dan lebih bersih, dan jika kalian tidak menemukan apa Anda bersedekah, Allah akan mengampuni Anda untuk pembicaraan khusus yang akan Anda berikan (sekali lagi, Allah Maha Penyayang untuk Anda). Maksudnya adalah bahwa berbicara khusus itu meskipun Anda tidak memberi sedekah bukanlah dosa". Di dalam ayat tersebut kita diajarkan untuk bermusyawarah untuk menentukan biaya pendidikan yang direncanakan, oleh karena itu wali murid, komite dan pihak sekolah harus melakukan musyawarah untuk merencanakan biaya pendidikan supaya semua berjalan dengan lancar dan ikhlas di hati para wali murid.

Berdasarkan ijma' dan sahabat menyimpulkan bahwa Negara memiliki kewajiban menangani pembiayaan pendidikan. Pada saat Khalifah Umar dan Utsman rela memberikan gajinya pada guru, Muadzin dan Imam Sholat Jama'ah, dan pada Khalifah Umar memberikan gaji tersebut dari pendapatan negara yang bersumber dari jizyah, kharaj, dan usyur.

Dalam tafsir Al-Lubāb fi al-'Ulūm Al-qur'an, permasalahan sumber daya manusia tidak bisa dikatakan sebagai memadai, ditambah dengan terbatasnya pengetahuan tentang sirkulasi dan pengaturan anggaran pembiayaan pendidikan, pengaturan anggaran yang tidak baik akan berdampak pada kurangnya analisis jangka panjang tentang bagaimana, mengapa, dalam hal ini mengakibatkan tidak ada upaya untuk mencari dan mengembangkan peluang. Pada akhirnya, lembaga

atau sekolah tersebut bersifat eksklusif dan hanya bisa mengandalkan dana pemerintah.

Surah Ali Imrān ayat: 14 :

رُئِيَ لِلنَّاسِ حُبُّ الشَّهَوَاتِ مِنَ النِّسَاءِ وَالْبَنِينَ وَالْقَنَاطِيرِ الْمُقَنْطَرَةِ مِنَ الذَّهَبِ وَالْفِضَّةِ وَالْخَيْلِ
(١٤) الْمُسَوَّمَةِ وَالْأَنْعَامِ وَالْحَرْثِ ۗ ذَلِكَ مَتَاعُ الْحَيَاةِ الدُّنْيَا ۗ وَاللَّهُ عِنْدَهُ حُسْنُ الْمَاٰبِ

Artinya : Dijadikan indah pada (pandangan) manusia kecintaan kepada apa-apa yang diingini, yaitu: wanita-wanita, anak-anak, harta yang banyak dari jenis emas, perak, kuda pilihan, binatang-binatang ternak dan sawah ladang. Itulah kesenangan hidup di dunia, dan di sisi Allah-lah tempat kembali yang baik (surga). (QS. Ali Imrān :14)

Ayat tersebut di tafsirkan dari tafsir Al- Jalalain yakni “(Dijadikan indah pada pandangan manusia kecintaan kepada syahwat) yakni segala yang disenangi serta diingini nafsu sebagai cobaan dari Allah atau tipu daya dari setan (yaitu wanita-wanita, anak-anak dan harta yang banyak) yang berlimpah dan telah berkumpul (berupa emas, perak, kuda-kuda yang tampan) atau baik (binatang ternak) yakni sapi dan kambing (dan sawah ladang) atau tanam-tanaman. (Demikian itu) yakni yang telah disebutkan tadi (merupakan kesenangan hidup dunia) di dunia manusia hidup bersenang-senang dengan hartanya, tetapi kemudian lenyap atau pergi (dan di sisi Allahlah tempat kembali yang baik) yakni surga, sehingga itulah yang seharusnya menjadi idaman dan bukan lainnya.

Pada ayat tersebut di sampaikan, ketika Allah menceritakan kondisi orang-orang beriman dengan do’a mereka agar diteguhkan dalam keimanan, kemudian di ceritakan bagaimana cerita orang kafir, dan yang menyebabkan kekufuran mereka, karena terpedaya dengan kehidupan dunia, banyaknya harta, dan keturunan (Yaqin et al., 2021). Ayat ini mengajarkan kita untuk beriman kepada Allah SWT dengan menginfakkan harta kita di jalan kebaikan yang di ridhoi Allah SWT. Yang kita niatkan dengan niat ikhlas, harta kita untuk anak-anak kita belajar di sekolah supaya bisa menjadi ladang pahala menuju surganya.

Oleh karena itu, para orang tua/ wali murid dan pemerintah bertanggung jawab terhadap pendidikan anak-anak, diantaranya adalah perintah untuk membaca dan belajar bagi individu maupun penduduk suatu negara. orang tua

wali murid dan pemerintah juga bertanggung jawab terhadap suksesnya kegiatan pendidikan ini. Agar pendidikan berjalan dengan lancar maka perlu disertai dengan biaya pendidikan. Hal tersebut tersirat pada surat Al- Alaq ayat 1-5:

إِقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ۚ اقْرَأْ وَرَبُّكَ الْأَكْرَمُ الَّذِي عَلَّمَ بِالْقَلَمِ ۗ عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ

Artinya : Dengan menyebut nama Tuhan Penciptamu, bacalah. Dia menjadikan manusia dari segumpal darah. Bacalah dengan teliti, dan Penguasamu adalah Yang Maha Bijaksana, Yang menginstruksikan (manusia) melalui firman, Dia menunjukkan kepada manusia apa yang tidak dia ketahui”.

Sesuai dengan tafsir Al- Jalalain yakni “(Bacalah) maksudnya mulailah membaca dan memulainya (dengan menyebut nama Rabbmu yang menciptakan) semua makhluk. (Dia telah menciptakan manusia) atau jenis manusia (dari 'alaq) lafal 'Alaq bentuk jamak dari lafal 'Alaqah, artinya segumpal darah yang kental. (Bacalah) lafal ayat ini mengukuhkan makna lafal pertama yang sama (dan Rabbmulah Yang Paling Pemurah) artinya tiada seorang pun yang dapat menandingi kemurahan-Nya. Lafal ayat ini sebagai Haal dari Dhamir yang terkandung di dalam lafal Iqra'. (Yang mengajar) manusia menulis (dengan qalam) orang pertama yang menulis dengan memakai qalam atau pena ialah Nabi Idris a.s. (Dia mengajarkan kepada manusia) atau jenis manusia (apa yang tidak diketahuinya) yaitu sebelum Dia mengajarkan kepadanya hidayah, menulis dan berkreasi serta hal-hal lainnya. Pada ayat tersebut mengajak dan memerintahkan belajar, menulis, dan membaca karena ketiganya adalah syiar Islam.

Menurut hadits Nabi Muhammad SAW, kewajiban orang tua untuk membiayai pendidikan anak merupakan sumber untuk membayar biaya pendidikan, yang juga dapat dipahami sebagai berikut.

“Hak seorang anak atas orang tuanya adalah memberinya nama baik dan budi pekerti yang baik, mengajarnya menulis, berenang, dan memanah, memberinya rejeki saja, dan menikahkannya jika sudah menemukan jodohnya.” (HR. al-Hakim)

Secara keseluruhan hadis-hadis tersebut menekankan bahwa orang tua memiliki tanggung jawab untuk mendidik anak-anaknya agar dapat mengembangkan akhlak yang baik, mengamalkan ibadah, membaca Al-Qur'an, dan menjadi anak yang sholeh. Biaya pendidikan anak harus ditanggung oleh orang tua anak jika anak ingin memiliki pendidikan seperti itu. Dana yang paling stabil adalah yang berasal dari sumbangan siswa (wali murid) dan mendukung proses pendidikan di sekolah.

Pada penelitian (Iskandar, 2019) di hasilkan akurasi 95,97% dan presisi 94,96%, yang berarti algoritma Decision Tree memiliki korelasi yang baik dalam penelitian prediksi tunggakan pembayaran uang sekolah.

Kemudian penelitian oleh (Muqorobin et al., 2019) hasil penerapan gabungan antara metode algoritma Information Gain dengan Naïve Bayes untuk prediksi keterlambatan pembayaran biaya pendidikan diperoleh hasil akurasi lebih baik yaitu 90%. Sedangkan jika tanpa penerapan algoritma information gain hanya diperoleh akurasi 80%.

Permasalahan tersebut menjadi landasan dalam penyusunan thesis ini, yang berjudul “Prediksi Kondisi Pembayaran Biaya Pendidikan Sekolah Berbasis Klasifikasi Ekonomi Siswa”, dengan tujuan melakukan prediksi kondisi pembayaran biaya pendidikan dengan mengklasifikasi ekonomi siswa dan melakukan perbandingan tingkat performansi yang dihasilkan oleh model data mining, sehingga diperoleh metode manakah yang memiliki tingkat akurasi yang lebih baik untuk memprediksi kondisi pembayaran dan mengklasifikasi ekonomi siswa, pengolahan berdasarkan data atribut: penghasilan, tanggungan keluarga, pekerjaan Orang tua. Dengan hasil prediksi tersebut dapat dibuat sebagai solusi dalam menangani permasalahan kondisi pembayaran biaya pendidikan dan pihak sekolah dapat melakukan tindakan berupa pembinaan terhadap siswa-siswi yang diperkirakan akan menunggak dalam melakukan pembayaran biaya pendidikan.

1.1.Pernyataan Masalah

Berdasarkan latar belakang masalah di atas maka permasalahan penelitian ini adalah

1. Metode klasifikasi mana yang paling optimal dalam memprediksi kondisi pembayaran biaya pendidikan sekolah?
2. Faktor- faktor apa yang berpengaruh dalam kinerja klasifikasi ekonomi siswa?

1.2.Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mengevaluasi metode klasifikasi dalam memprediksi kondisi pembayaran biaya pendidikan sekolah.
2. Menganalisis faktor- faktor yang berpengaruh dalam mengklasifikasi ekonomi siswa.

1.3.Manfaat Penelitian

Hasil dari penelitian ini dapat digunakan oleh para pihak sekolah dasar dan sekolah menengah pertama dalam menangani biaya operasional yang bersumber dari siswa.

1.4. Batasan Masalah

Adapun batasan masalah adalah sebagai berikut :

1. Data yang digunakan diperoleh dari data kondisi siswa di MI Miftahul Huda Sukolilo.
2. Data siswa yang diambil Tahun Ajaran 2016/ 2017 sampai dengan 2021/2022

1.5.Sistematika Penulisan

Sistematika penulisan diperlukan untuk memahami isi dari penulisan thesis. Berikut adalah uraian singkat yang dibahas pada masing-masing bab:

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang masalah, rumusan masalah, , tujuan dan manfaat penelitian baik bagi obyek, ilmu, dan penyusun thesis serta ruang lingkup penelitian. Selain hal itu bab ini juga menjelaskan tentang sistematika penyusunan laporan.

BAB II STUDI PUSTAKA

Bab ini menerangkan tentang studi pustaka dari laporan-laporan yang telah ada sebelumnya dan teori-teori tentang prediksi kondisi pembayaran biaya pendidikan sekolah berbasis klasifikasi ekonomi siswa.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang metodologi yang digunakan dalam penelitian berupa metode pengumpulan data dan analisis serta kerangka berfikir penulis dalam lakukan beberapa penelitian. ini.

BAB IV PEMBAHASAN

Bab ini membahas tentang pembahasan terhadap implementasi metode klasifikasi dan hasil klasifikasi.

BAB V KESIMPULAN

Bab ini merupakan bab terakhir yang berisi tentang kesimpulan serta saran dari penyusun.

BAB II

STUDI PUSTAKA

2.1.Sistem Prediksi Biaya Pendidikan Sekolah

Penelitian terkait mengenai pembahasan prediksi kondisi pembayaran biaya pendidikan sekolah berbasis klasifikasi ekonomi siswa telah banyak dilakukan pada penelitian sebelumnya.

Menurut Bahtiar & Firmansyah (2017) permasalahan yang muncul terkait pembayaran *SPP* yaitu apabila para santri terlambat membayarkan *SPP* dari waktu yang telah ditentukan. Hal ini menjadi masalah karena pembayaran *SPP* adalah salah satu faktor penting untuk memberikan kualitas pelayanan yang baik bagi pihak pondok. Oleh karena itu di perlukan implementasi data mining, pada fase ini dilakukan analisis data dengan menggunakan Algoritma Naive Bayes Classifier. Sehingga menghasilkan hasil pengujian akurasi 45 data training dan 22 data testing dari model yang dihasilkan diperoleh tingkat accuracy sebesar 63,64%.

Hasil penelitian dari Apandi *et al.* (2019) data yang digunakan adalah data survei dalam bentuk kuesioner yang dibagikan kepada mahasiswa program studi teknik informatika Politeknik TEDC Bandung yang terdiri dari beberapa aspek atau faktor penilaian. Untuk membuat klasifikasi mahasiswa yang terlambat dan tepat waktu dalam melakukan pembayaran *SPP* di Politeknik TEDC menggunakan metode Algoritma C4.5 dengan mendapatkan hasil tingkat akurasi tertinggi pada data partisi 2 sebanyak 80% data training dan menghasilkan akurasi sebesar 75,00%.

Selanjutnya penelitian oleh Muqorobin *et al.* (2019) dengan data training dan data testing menggunakan data kelengkapan siswa (nomor induk, nama siswa, kelas, alamat, pendapatan orang tua, pendidikan, tanggungan keluarga dan usia) pada tahun ajaran 2017/2018. Hasil penerapan gabungan antara metode algoritma Information Gain dengan Naïve Bayes untuk prediksi keterlambatan pembayaran

biaya pendidikan diperoleh hasil akurasi lebih baik yaitu 90%. Sedangkan jika tanpa penerapan algoritma information gain hanya diperoleh akurasi 80%.

Sedangkan penelitian oleh Ginting *et al.* (2020) Penelitian tersebut dilakukan dengan menggunakan beberapa variabel seperti jumlah pendapatan, tanggungan keluarga, latar belakang pendidikan orang tua dan usia orang tua. Dataset yang digunakan pada penelitian ini adalah dari penelitian yang sudah dilakukan oleh Muqorobin, 2019 dan diimplementasikan kedalam bentuk pemrograman menggunakan bahasa pemrograman python untuk menghasilkan keterangan hasil prediksi. Hasil penelitian yang didapatkan mendapatkan tingkat akurasi sebesar 73%.

Hasil penelitian Prabowo *et al.* (2021) Pembayaran *SPP* menjadi masalah penting di pondok pesantren Al – Arifah karena banyak dari para *santri* atau orang yang belajar di pesantren terlambat membayar *SPP* meskipun sudah ditetapkan batas waktu pembayaran. Untuk mengatasi masalah ini, dengan menggunakan data mining, perlu untuk mengidentifikasi faktor-faktor yang menunda pembayaran uang sekolah. Dengan klasifikasi menggunakan algoritma Naive Bayes dihasilkan akurasi sebesar 64.64%.

Menurut Purba *et al.* (2019) dalam mengumpulkan data penulis menyebarkan kuesioner kepada mahasiswa di STIKOM Tunas Bangsa Pematangsiantar. C1 (Penghasilan Orang Tua), C2 (Jumlah Tanggungan), C3 (Pekerjaan Orang Tua), C4 (Status Tempat Tinggal), C5 (Penyalahgunaan Uang), dan C6 (Faktor Eksternal) adalah parameter yang digunakan. Kesimpulan penelitian adalah algoritma Naive Bayesian Classification (NBC) dapat digunakan untuk mengklasifikasikan faktor penyebab tunggakan *SPP*. Hasil tes tersebut akurat hingga 62,63 persen, dengan 39 responden menyatakan tidak disiplin karena tidak membayar *SPP*, dan 116 responden menyatakan disiplin atau membayar *SPP* tepat waktu. Prediksi No memiliki nilai 60,34 persen dengan presisi kelas, sedangkan prediksi Yes memiliki nilai 69,03 persen. Class recall memiliki nilai 85,37 persen untuk benar Ya dan 36,99% untuk benar Tidak. Dengan kata lain, variabel C1 (Pendapatan Orang Tua), C2 (Jumlah Tanggungan), C3 (Pekerjaan

Orang Tua), dan C4 (Eksternal) Factors) merupakan empat faktor yang menyebabkan mahasiswa tertinggal dalam pembayaran SPP. Komponen terbesar adalah variabel luar dengan konsekuensi 0,804.

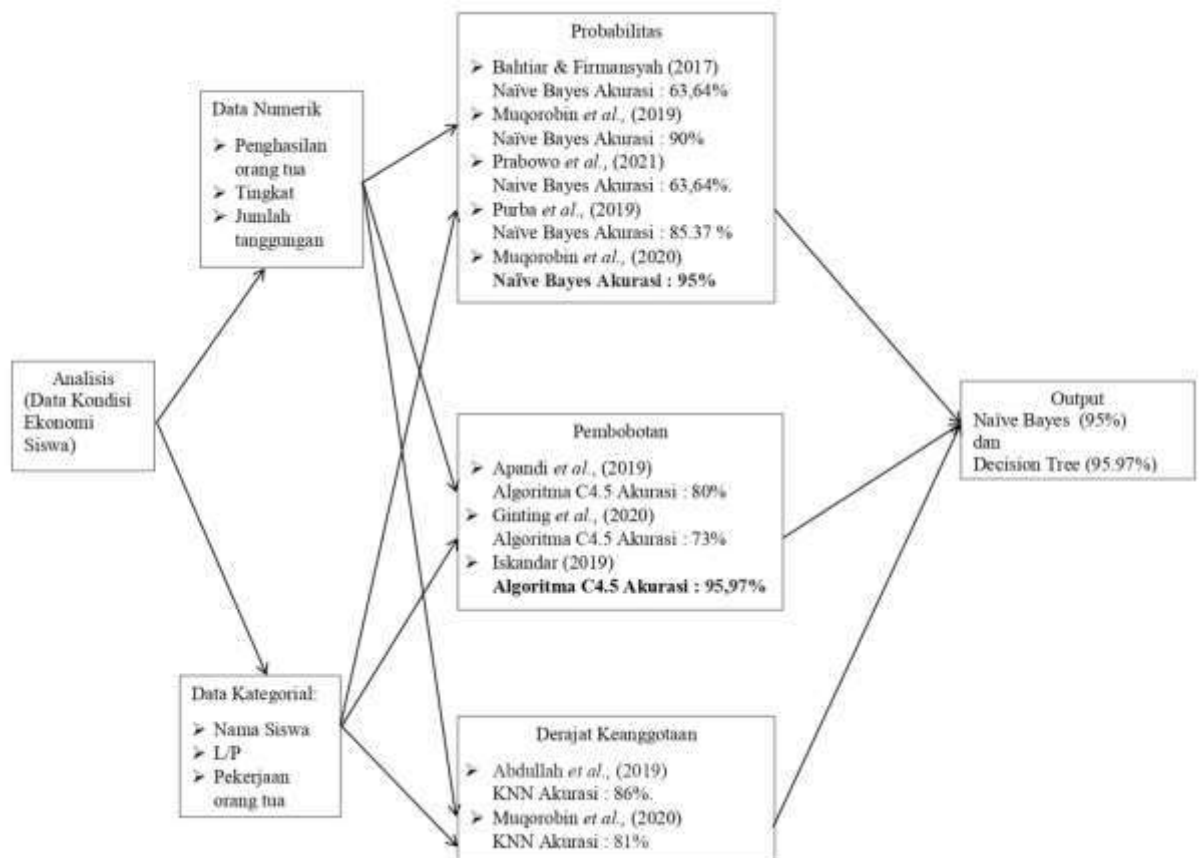
Penelitian dari Muqorobin *et al.* (2020) data yang digunakan dalam penelitian ini bersumber dari Data Berbasis Pendidikan sekolah dinas tahun pelajaran 2017/2018 dan laporan transaksi pembayaran SPP Sekolah. Atribut data yang akan digunakan sebagai variabel determinan adalah Pendapatan Orang Tua, Tanggungan Keluarga, Pendidikan Ayah, Usia, Pendidikan Ibu dan Usia Ibu. Dari hasil membandingkan tingkat akurasi metode Naïve Bayes dengan metode K-Nearest Neighbor diperoleh Metode Naïve Bayes + Metode Information Gain menghasilkan akurasi tertinggi yaitu 95% dibandingkan dengan metode Naïve Bayes saja yaitu 85% dan metode K-NN yaitu 81%.

Menurut Iskandar (2019) data yang diambil dari dua database master yaitu data iuran sekolah dari SMK Angkasa, dan besaran gaji orang tua dikumpulkan dari data dasar pendidikan SMK Angkasa Tasikmalaya yang kemudian dilakukan analisis pada data tersebut, yang menghasilkan akurasi 95,97% dan presisi 94,96%, yang berarti algoritma Decision Tree memiliki korelasi yang baik dalam penelitian prediksi tunggakan pembayaran uang sekolah.

Kemudian pada penelitian Abdullah *et al.* (2019) data didapatkan dari hasil wawancara pada bagian administrasi guna untuk mengetahui variable yang mempengaruhi dalam melakukan prediksi keterlambatan pembayaran SPP. Maka dari hasil wawancara tersebut, diperoleh data yang digunakan untuk memprediksi, yaitu data siswa dan data pembayaran SPP di SMK Al-Islam Surakarta. Dari hasil prediksi menggunakan metode k-nearest neighbor, dihasilkan Akurasi sebesar 86%.

2.2. Kerangka Teori

Bahasan kerangka teori pada penelitian ini yaitu merujuk dari penelitian terdahulu atau penelitian sebelumnya yang relevan sebagai referensi. Berikut ini hasil tahapan metode yang merujuk dari penelitian yang relevan sesuai dengan bahasan penelitian ini. Kerangka teori ditunjukkan pada Gambar 2.1.



Gambar 2. 1 Kerangka teori

Kerangka teori dijelaskan melalui langkah- langkah yang dilakukan sebagai berikut :

- Data yang digunakan diperoleh dari data kondisi siswa di MI Miftahul Huda Sukolilo
- Adapun jurnal yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2.1.

Tabel 2. 1 Daftar Jurnal

No	Nama (tahun)	Judul	Subjek Penelitian	Metode	Hasil Akurasi
1	1. Agus Bahtiar 2. Mulyawan 3. Suryani 4. Dindin Firmansyah (2017)	Klasifikasi Ketepatan Waktu Pembayaran SPP Di Pondok Pesantren Al-Arifah Menggunakan Algoritma Naive Bayes	Pondok Pesantren Al – Arifah	Algoritma Naive Bayes	Diperoleh tingkat accuracy Sebesar 63,64%
2	1. Tri Herdiawan Apandi 2. Roby Bayu Maulana 3. Rian Piarna 4. Dwi Vernanda (2019)	Menganalisis Kemungkinan Keterlambatan Pembayaran SPP Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung)	Politeknik Tedc	Algoritma C4.5	Diperoleh tingkat akurasi tertinggi pada data partisi 2 sebanyak 80% data training dan menghasilkan akurasi sebesar 75,00%.
3	1. Muqorobin 2. Kusrini 3. Emha Taufiq Luthfi (2019)	Optimasi Metode Naïve Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah	Data kelengkapan siswa (nomor induk, nama siswa, kelas, alamat, pendapatan orang tua, pendidikan, tanggungan keluarga dan usia) pada tahun ajaran 2017/2018 di Smk Al-Islam Surakarta	Metode Naive Bayes	Penerapan algoritma information gain menghasilkan tingkat akurasi 90%, sedangkan tingkat akurasi tanpa algoritma hanya 80%.
4	1. Victor Saputra Ginting 2. Kusrini 3. Emha Taufiq (2020)	Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python	Smk Al-Islam Surakarta dengan variabel seperti jumlah pendapatan, tanggungan keluarga, latar belakang pendidikan orang tua dan usia orang tua.	Algoritma C4.5	Mencapai tingkat akurasi 90%.
5	1. Candi Prabowo 2. Lukman Nul Hakim 3. Cep Lukman Rohmat 4. Ade Rizki Rinaldi (2021)	Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran	Pondok Pesantren Al – Arifah (Buntet Pesantren Cirebon)	Algoritma <i>Naive Bayes</i> .	Model yang dihasilkan diperoleh tingkat <i>accuracy</i> sebesar 63,64%.

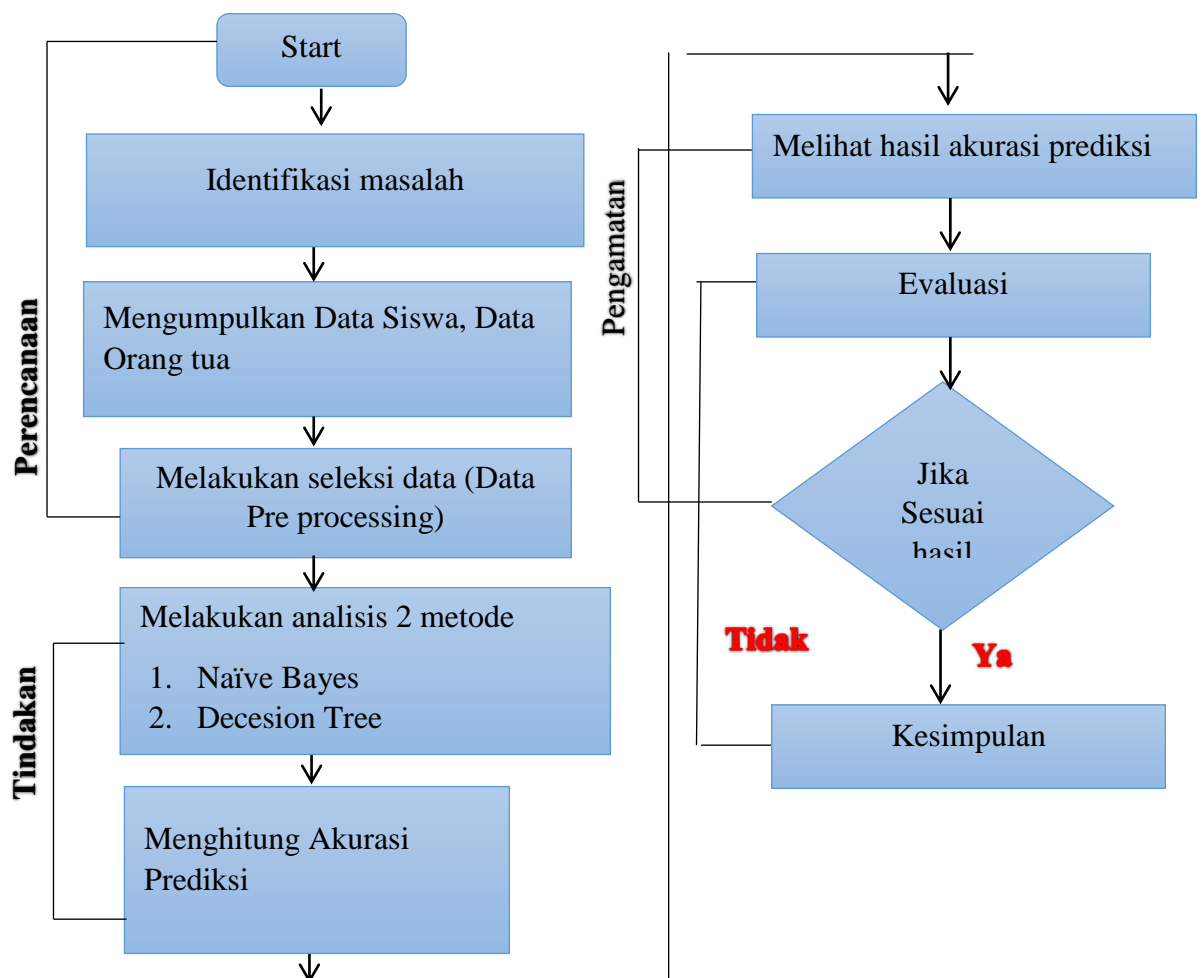
No	Nama (tahun)	Judul	Subjek Penelitian	Metode	Hasil Akurasi
6	1. Ledis Pebriani Purba 2. Dedy Hartama 3. Eka irawan 4. Anjar Wanto (2019)	Memprediksi Faktor Tunggakan Uang Kuliah Menggunakan Metode Naive Bayes	Data Mahasiswa di STIKOM Tunas Bangsa: C1 (pendapatan orang tua), C2 (pekerjaan tanggungan), C3 (pekerjaan orang tua), C4 (status tempat tinggal), C5 (penyalahgunaan uang), dan C6 (faktor eksternal)	Algoritma <i>Naive Bayes</i> .	Model yang dihasilkan diperoleh tingkat <i>accuracy</i> sebesar 85.37 %.
7	1. Muqorobin 2. Kusrini 3. Emha Taufiq Luthfi (2020)	Estimation System For Late Payment Of School Tuition Fees	Data Berbasis Pendidikan sekolah dinas tahun pelajaran 2017/2018 dan laporan transaksi pembayaran SPP Sekolah. Atribut data yang akan digunakan sebagai variabel determinan adalah Pendapatan Orang Tua, Tanggungan Keluarga, Pendidikan Ayah, Usia, Pendidikan Ibu dan Usia Ibu.	Algoritma <i>Naive Bayes</i> . Algoritma KNN	Model yang dihasilkan diperoleh tingkat <i>accuracy</i> dengan menggunakan Metode Naive Bayes 95% , dan dengan metode K-NN yaitu 81%.
8	Iskandar (2019)	Parents' Sum of Salaries Analyses towards School Tuition Fee Arrears Potential with Decision Tree Method	Data iuran sekolah dari SMK Angkasa dan besaran gaji orang tua	Algoritma <i>Decision Tree</i>	Model yang dihasilkan diperoleh tingkat <i>accuracy</i> sebesar 95,97%
9	1. Robi Wariyanto Abdullah 2. Kusrini 3. Emha Taufiq Luthfii (2019)	Prediksi Keterlambatan Pembayaran Spp Sekolah Dengan Metode K-Nearest Neighbor (Studi Kasus Smk Al-Islam Surakarta)	Data siswa dan data pembayaran SPP di SMK Al-Islam Surakarta	Algoritma K-Nearest Neighbor	Model yang dihasilkan diperoleh tingkat <i>accuracy</i> sebesar 86%

Dari hasil rujukan penelitian terdahulu yang di uraikan di atas, maka di gunakan metode *Naïve Bayes* dan *Decision Tree*, dengan pengembangan aplikasi menggunakan *Python* sebagai alat bantu analisis.

BAB III DESAIN PENELITIAN

3.1. Desain Penelitian

Metode analisis data menggunakan *Naïve Bayes* dan *Decesion Tree* yang merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan asumsi ketidaktergantungan yang tinggi. Keuntungan menggunakan *Naïve Bayes* dan *Decesion Tree* adalah metode ini hanya membutuhkan jumlah data pelatihan (data training) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Berikut ini adalah desain penelitian dengan metode *Naïve Bayes* dan *Decision Tree*, dapat ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Desain Penelitian

3.1.1. Perencanaan

1. Pada tahap awal yaitu tahap perencanaan, peneliti melakukan identifikasi masalah dengan menentukan latar belakang masalah, mencari rumusan masalah, parameter dan solusi permasalahan.
2. Data yang digunakan diperoleh dari data kondisi siswa di MI Miftahul Huda Sukolilo. Data siswa yang di ambil Tahun Ajaran 2016/ 2017 sampai dengan 2021/2022 ditampilkan pada Tabel 3.1. Untuk mengumpulkan informasi yang berkaitan dengan studi kasus penelitian.

Tabel 3. 1 Data siswa

No	Kode Sistem	No Induk	Nisn	Nama	Kelamin	Tempat Lahir	Tgl Lahir (Yyyy-Mm-Dd)	Alamat	Tingkat	Tahun Masuk
1	2051824523345a10cc01	1.11235e+17		A'an Krono Aji	L	Malang	1/12/2008	Gandon Barat Rt. 15 Rw. 02 Sukolilo Jabung	6	2016
2	2051824527d0b38bd002	1.11235e+17		Aan Rian Pratama	L	Malang	2/25/2011	Jamparing Pakis	4	2018
3	205182452c7297a0a401	1.11235e+17		Abdillah Dwi Kurnia Santoso	L	Malang	3/11/2013	Kampung Anyar	2	2020
4	20518245257820738f01	1.11235e+17		Abellita Yuniar Syaputri	P	Malang	6/18/2010	Kp. Anyar Rt. 08 Rw. 01 Sukolilo Jabung	5	2017
5	2051824523345a10cc02	1.11235e+17		Achmad Arifin	L	Malang	4/10/2009	Jl. Gatot Subroto Rt. 04 Rw. 03 Kemantran	6	2016
6	205182452a197b410c01	1.11235e+17		Achmad Fawaidul Ilmi	L	Malang	10/25/2011	Perumnas Jabung Permai Kemantran	3	2019
7	205182452ec1ee43bb01	1.11235e+17	82113654	Achmad Galih Permana	L	Batam	4/29/2008	Tebelo Sidomulyo	6	2016

No	Kode Sistem	No Induk	Nisn	Nama	Kelamin	Tempat Lahir	Tgl Lahir (Yyyy-Mm-Dd)	Alamat	Tingkat	Tahun Masuk
8	2051824 527d0b3 8bd003	1.1123 5e+17		Achmad Revi Firmansyah	L	Malang	6/4/2010	Gandon Barat Sukolilo	4	2018
9	2051824 52a197b 410c02	1.1123 5e+17		Addini Nafisatur Rohmah	P	Malang	4/3/2012	Kampung Anyar	3	2019
10	2051824 5230628 77f901	1.1123 5e+17		Adelia Zahira	P	Malang	12/17/2007	Jl. Raya Sukolilo	6	2016
...
418	2051824 527d0b3 8bd049	1.1123 5e+17		Zahrotul Sarirotul Mukaromah	P	Malang	9/25/2011	Sukolilo	4	2018
419	2.05e+1 9	1.1123 5e+17		Zaki Ahmad Ramadhani	L	Malang	9/15/2009	Sukolilo	5	2017
420	2051824 527d0b3 8bd04a	1.1123 5e+17		Zaky Syahbana Putra Wicaksono	L	Malang	7/4/2012	Jl. Dewi Sartika Rt.04 Rw.04 Kematren	4	2018

Dari tabel 3.2. dapat dilihat bahwa data siswa memiliki 10 atribut yaitu:

Tabel 3. 2 Tabel data siswa

Atribut	Kriteria
Kode Sistem	Memiliki
No Induk	Memiliki Tidak memiliki
NISN,	Memiliki Tidak memiliki
Nama,	Memiliki
Kelamin	Memiliki
Tempat Lahir,	Memiliki
Tgl lahir (yyyy-mm-dd),	Memiliki
Alamat,	Memiliki
Tingkat,	Memiliki
Tahun Masuk	Memiliki

Pada Tabel 3.1. disebutkan bahwa data siswa memiliki 10 atribut, kode sistem merupakan kode sistem siswa yang dimiliki siswa dari sistem pemutakhiran data EMIS, 2051824523345A10CC01 sampai dengan 2051824527D0B38BD04A. Atribut no induk didapatkan siswa dari sekolah, waktu awal masuk sekolah. Atribut NISN merupakan nomor induk sekolah nasional yang diperoleh siswa dari Diknas, NISN ini diperoleh siswa mulai sekolah di TK, jika ada siswa yang tidak menempuh sekolah TK maka pihak sekolah yang akan mendaftarkannya. Atribut nama merupakan atribut penting untuk identitas siswa. Atribut kelamin merupakan data yang mengategorikan jenis kelamin para siswa laki-laki atau perempuan. Atribut tempat tanggal lahir, merupakan atribut yang berisi tempat dimana siswa itu di lahirkan. Data ini diperoleh dari kutipan akta kelahiran siswa. Sedangkan atribut alamat, merupakan data domisili siswa sesuai dengan kartu keluarga (KK). Untuk atribut tingkat merupakan data yang menandakan siswa berada pada tingkatan kelas berapa, 1,2,3,4,5, dan 6. Sedangkan tahun masuk merupakan tahun dimana siswa ini pertama kali masuk di sekolah. Data ini bisa di lihat pada buku induk siswa dan ditampilkan pada Tabel 3.3.

Tabel 3. 3 Data orang tua

No	Nama	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan
1	A'an Krono Aji	Swasta	1,500,000	2
2	Aan Rian Pratama	Swasta	1,500,000	1
3	Abdillah Dwi Kurnia Santoso	Petani	500,000	1
4	Abellita Yuniar Syaputri	Tni	3,000,000	4
5	Achmad Arifin	Tni	3,000,000	5
6	Achmad Fawaidul Ilmi	Guru	500,000	2
7	Achmad Galih Permana	Swasta	1,500,000	3
8	Achmad Revi Firmansyah	Swasta	1,500,000	7
9	Addini Nafisatur Rohmah	Swasta	1,500,000	2
10	Adelia Zahira	Swasta	1,500,000	1
..
418	Zahrotul Sarirotul Mukaromah	Wirasaha	3,000,000	2
419	Zaki Ahmad Ramadhani	Swasta	500,000	2
420	Zaky Syahbana Putra Wicaksono	Swasta	1,500,000	3

Dari tabel 3.4. dapat dilihat bahwa data siswa memiliki 4 atribut yaitu:

Tabel 3. 4 Tabel Atribut data orang tua

Atribut	Kriteria
Nama,	Memiliki
Pekerjaan orang tua	Memiliki
Penghasilan orang tua per bulan	Memiliki
Jumlah tanggungan	Memiliki

Pada Tabel 3.4.. disebutkan bahwa data orang tua memiliki 5 atribut, atribut nama merupakan atribut penting untuk identitas siswa, nama yang di munculkan adalah nama siswa. Atribut pekerjaan orang tua, atribut ini berisi macam- macam pekerjaan yang di miliki oleh orang tua, yaitu swasta, pegawai negeri, guru, TNI, dan wiraswasta. Pada atribut penghasilan orang tua perbulan merupakan atribut yang berisi data penghasilan orang tua mulai dari jumlah gaji Rp. 500.000 – Rp. 5.000.000. Tanggungan orang tua adalah jumlah tanggungan anak yang masih di biayai orang tua.

3. Selanjutnya data yang sudah di kumpulkan merupakan data yang belum terstruktur, maka membutuhkan *pre-processing*, *pre-processing* dilakukan pada proses data *training* melalui :

a. *Cleansing*

Pada proses ini, data yang diperoleh dibersihkan dari atribut yang kosong, menghapus data yang kosong, melakukan penghapusan data yang kurang penting dan hanya mengambil data siswa dan atribut dari data orang tua siswa yang berguna sebagai parameter prediksi pembayaran biaya pendidikan pendidikan. Data siswa yang dihapus adalah atribut no, nama, kode sistem, no induk, nisn, tempat lahir, tgl lahir (yyyy-mm-dd), alamat, tingkat, tahun masuk. Dimana data siswa yang sudah dicleaning terlihat di Tabel 3.5.

Tabel 3. 5 Data siswa sudah di normalisasi

No	Nama	Kelamin
1	A'an Krono Aji	L
2	Aan Rian Pratama	L

No	Nama	Kelamin
3	Abdillah Dwi Kurnia Santoso	L
4	Abellita Yuniar Syaputri	L
5	Achmad Arifin	L
6	Achmad Fawaidul Ilmi	L
7	Achmad Galih Permana	L
8	Achmad Revi Firmansyah	P
9	Addini Nafisatur Rohmah	P
10	Adelia Zahira	P
...
418	Zahrotul Sarirotul Mukaromah	P
419	Zaki Ahmad Ramadhani	L
420	Zaky Syahbana Putra Wicaksono	L

Data orang tua yang dihapus adalah atribut nama. Dimana data orang tua yang sudah dicleaning dapat dilihat pada Tabel 3.6.

Tabel 3. 6 Data orang tua sudah di normalisasi

No	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan
1	Swasta	1,500,000	2
2	Swasta	1,500,000	1
3	Petani	500,000	1
4	Tni	3,000,000	4
5	Tni	3,000,000	5
6	Guru	500,000	2
7	Swasta	1,500,000	3
8	Swasta	1,500,000	7
9	Swasta	1,500,000	2
10	Swasta	1,500,000	1
..
418	Wirausaha	3,000,000	2
419	Swasta	500,000	2
420	Swasta	1,500,000	3

b. Data Integrasi

Data yang telah bersih dari missing value dan perulangan data selanjutnya digabungkan kedalam satu tabel utama yang dapat digunakan sebagai data akhir data kondisi siswa. Data yang sudah diintegrasikan akan dilakukan seleksi atribut untuk memilih data yang relevan sesuai pada kebutuhan yang akan dicapai. Adapun data hasil integrasi pada Tabel 3.7:

Tabel 3. 7 Data sudah di normalisasi

No	Nama	Kelamin	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan
1	A'an Krono Aji	L	Swasta	1,500,000	2
2	Aan Rian Pratama	L	Swasta	1,500,000	1
3	Abdillah Dwi Kurnia Santoso	L	Petani	500,000	1
4	Abellita Yuniar Syaputri	L	Tni	3,000,000	4
5	Achmad Arifin	L	Tni	3,000,000	5
6	Achmad Fawaidul Ilmi	L	Guru	500,000	2
7	Achmad Galih Permana	L	Swasta	1,500,000	3
8	Achmad Revi Firmansyah	P	Swasta	1,500,000	7
9	Addini Nafisatur Rohmah	P	Swasta	1,500,000	2
10	Adelia Zahira	P	Swasta	1,500,000	1
...
418	Zahrotul Sarirotul Mukaromah	P	Wirausaha	3,000,000	2

419	Zaki Ahmad Ramadhani	L	Swasta	500,000	2
420	Zaky Syahbana Putra Wicaksono	L	Swasta	1,500,000	3

Tabel 3.7. merupakan tabel penggabungan antara tabel siswa dan tabel orang tua. Yang menjadi tabel kondisi siswa, dengan memiliki 5 atribut, yaitu : nama, kelamin, pekerjaan orang tua, penghasilan orang tua per bulan dan jumlah tanggungan orang tua.

c. Data Selection

Data –data kondisi siswa tahun ajaran 2016/ 2017 – 2021/ 2022. Untuk data untuk pengujian dan pelatihan yaitu data kondisi siswa yang terlambat, tepat dan tidak bayar. Data yang dibutuhkan yaitu data nama, kelamin, pekerjaan orang tua, penghasilan orang tua, dan tanggungan orang tua. Ditampilkan pada Tabel 3.8. yaitu :

Tabel 3. 8 Data selection

No	Nama	Kelamin	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan	Actual Class
1	A'an Krono Aji	L	Swasta	1,500,000	2	Terlambat
2	Aan Rian Pratama	L	Swasta	1,500,000	1	Tidak Bayar
3	Abdillah Dwi Kurnia Santoso	L	Petani	500,000	1	Terlambat
4	Abellita Yuniar Syaputri	L	Tni	3,000,000	4	Terlambat
5	Achmad Arifin	L	Tni	3,000,000	5	Tepat
6	Achmad Fawaidul Ilmi	L	Guru	500,000	2	Terlambat
7	Achmad Galih	L	Swasta	1,500,000	3	Terlambat

No	Nama	Kelamin	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan	Actual Class
	Permana					
8	Achmad Revi Firmansyah	P	Swasta	1,500,000	7	Tepat
9	Addini Nafisatur Rohmah	P	Swasta	1,500,000	2	Tidak Bayar
10	Adelia Zahira	P	Swasta	1,500,000	1	Terlambat
...
418	Zahrotul Sarirotul Mukaromah	P	Wirausaha	3,000,000	2	Tepat
419	Zaki Ahmad Ramadhani	L	Swasta	500,000	2	Terlambat
420	Zaky Syahbana Putra Wicaksono	L	Swasta	1,500,000	3	Tepat

d. Transformation

Pada tahapan ini yaitu melakukan konversi atribut yang bekerja. Tujuan dari konversi atribut yaitu merubah atribut bernilai kontinu (tidak terhingga) menjadi atribut dengan nilai nominal (berhingga). Terdapat dua cara untuk melakukan normalisasi data dengan membagi range data dari data yang terkecil sampai data yang digunakan. Berikut data kategori ditunjukkan pada Tabel 3.9.

Tabel 3. 9 Kategori kriteria gaji

Kriteria gaji		Parameter
500 – 1 juta	Cukup	C
1,1 - 2,5 juta	Cukup tinggi	CT
2,6 - 4 juta	Tinggi	T
4,1- 10 juta	Sangat tinggi	ST

Tabel kategori kriteria gaji di jelaskan bahwasannya gaji orang tua sebesar 500.000 – 1.000.000 masuk dalam kriteria gaji cukup yang di parameterkan

menjadi c, gaji orang tua sebesar 1.100.000 – 2.500.000 masuk dalam kriteria gaji cukup tinggi yang di parameterkan menjadi ct, gaji orang tua sebesar 2.600.000 – 4.000.000 masuk dalam kriteria gaji tinggi yang di parameterkan menjadi t, sedangkan gaji orang tua sebesar 4.100.000 – 10.000.000 masuk dalam kriteria gaji sangat tinggi yang di parameterkan menjadi st. Berikut data kategori ditunjukkan pada Tabel 3.10.

Tabel 3. 10 Kategori tanggungan orang tua

Kriteria jumlah tanggungan orang tua		Parameter
1- 2	Cukup	C
3- 10	Banyak	B

Tabel kategori tanggungan orang tua di jelaskan bahwasannya jumlah tanggungan anak yang dibiayai orang tua dengan jumlah 1-2 anak masuk dalam kategori jumlah cukup yang di parameterkan menjadi c, sedangkan jumlah tanggungan anak yang dibiayai orang tua dengan jumlah 3-10 anak masuk dalam kategori jumlah banyak yang di parameterkan menjadi b. Berikut data kategori ditunjukkan pada Tabel 3.11.

Tabel 3. 11 Kategori pekerjaan orang tua

Kriteria pekerjaan orang tua	Parameter
Swasta	S
Petani	P
TNI	T
Guru	G
Wirausaha	W
PNS	PN
Swasta	S

Tabel 3.11. menjelaskan macam- macam pekerjaan yang di miliki oleh orang tua, yaitu swasta di kategorikan s, pegawai negeri sipil dikategorikan s, guru dikategorikan g, TNI dikategorikan T, dan wiraswasta di kategorikan w.

Berikut adalah hasil dari kategori data kondisi siswa pada Tabel 3.1.1.

Tabel 3. 12 Data kondisi siswa

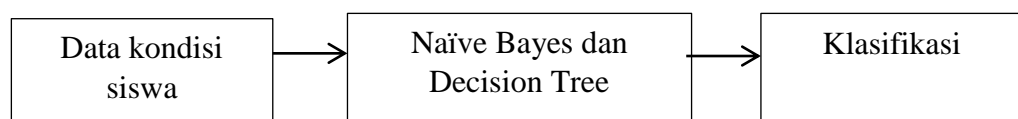
No	Nama	Kelamin	Pekerjaan Orang Tua	Penghasilan Orang Tua Per Bulan	Jumlah Tanggungan	Actual Class
1	A'an Krono Aji	L	P	Ct	C	Terlambat
2	Aan Rian Pratama	L	S	Ct	C	Tidak Bayar
3	Abdillah Dwi Kurnia Santoso	L	Pn	C	C	Terlambat
4	Abellita Yuniar Syaputri	L	T	T	C	Terlambat
5	Achmad Arifin	L	S	T	C	Tepat
6	Achmad Fawaidul Ilmi	L	G	C	C	Terlambat
7	Achmad Galih Permana	L	S	Ct	C	Terlambat
8	Achmad Revi Firmansyah	P	W	Ct	C	Tepat
9	Addini Nafisatur Rohmah	P	W	Ct	C	Tidak Bayar
10	Adelia Zahira	P	T	Ct	B	Terlambat
...
418	Zahrotul Sarirotul Mukaromah	P	W	T	B	Tepat
419	Zaki Ahmad Ramadhani	L	Pn	C	C	Terlambat
420	Zaky Syahbana Putra Wicaksono	L	S	Ct	C	Tepat

Penentuan sistem prediksi pada Tabel 3.12. dilakukan berdasarkan analisa pekerjaan orang tua, penghasilan orang tua perbulan dan jumlah tanggungan orang tua. Berdasarkan Tabel 3.12. diperoleh prediksi kondisi pembayaran biaya pendidikan sekolah terhadap klasifikasi ekonomi siswa, yang di mana jenis kelamin L, pekerjaan orang tua P, penghasilan orang tua perbulan CT dan tanggungan orang tua C, dia di prediksi membayar terlambat. Tabel 3.12. menjelaskan prediksi kondisi pembayaran biaya pendidikan sekolah terhadap

klasifikasi ekonomi siswa, yang di mana jenis kelamin L, pekerjaan orang tua S, penghasilan orang tua perbulan CT dan tanggungan orang tua C, dia di prediksi membayar tidak bayar.

3.1.2. Tindakan / Pelaksanaan

1. Pada tahap kedua yaitu tahap pelaksanaan peneliti memasukkan atribut. Berikut adalah desain sistem pada tahapan pelaksanaan proses prediksi pembayaran biaya pendidikan pendidikan.

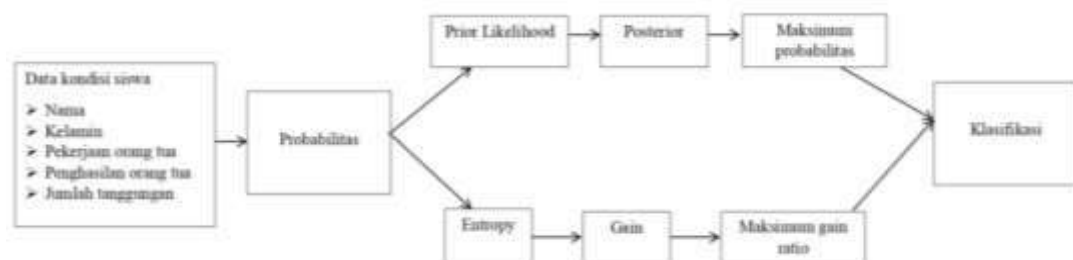


Gambar 3. 2 Desain Sistem

- a. Proses input data adalah memasukkan data kondisi siswa.
- b. Setelah melakukan input data, dilakukan proses prediksi kondisi pembayaran biaya pendidikan sekolah menggunakan metode *Naïve Bayes* dan metode *Decision Tree*.
- c. Setelah proses prediksi dilakukan, selanjutnya dilakukan klasifikasi ekonomi siswa.

3.1.3. Pengamatan

Tahap ke tiga adalah tahap pengamatan merupakan suatu tindakan untuk mengetahui kinerja system yang telah dibangun. Berikut konsep model sistem yang akan di bangun ditampilkan pada Gambar 3. 3.



Gambar 3. 3 Konsep model

Berdasarkan Gambar 3.3. tentang konsep model di jelaskan bahwa atribut atau parameter yang digunakan nantinya akan dilakukan proses pemberian actual class agar mempermudah dalam pengumpulan dataset.

Selanjutnya dilakukan perhitungan probabilitas pada masing-masing parameter, kemudian dapat diterapkan pada algoritma Naïve Bayes. Setelah probabilitas masing-masing parameter sudah di temukan, selanjutnya dilakukan menghitung prior dan likelihood sehingga akan menghasilkan nilai posterior. Nilai posterior tertinggi akan di pilih untuk diproses sehingga akan menghasilkan nilai prediksi actual class terlambat, tepat dan tidak bayar.

Proses selanjutnya menerapkan algoritma Decision Tree, setelah nilai probabilitas masing-masing parameter sudah diketahui, langkah selanjutnya menentukan berapa banyak kasus yang perlu diputuskan terlambat, jumlah hasil yang mungkin tepat, dan jumlah hasil yang mungkin tidak bayar, setelah itu dilakukan langkah menghitung entropy untuk semua kasus, menghitung entropy untuk sub kasus (berdasarkan atribut) dan menghitung gain untuk setiap atribut. Atribut dengan gain tertinggi dapat menjadi node akar. Node akar yang terbentuk akan di jadikan model dalam memprediksi actual class terlambat, tepat dan tidak bayar.

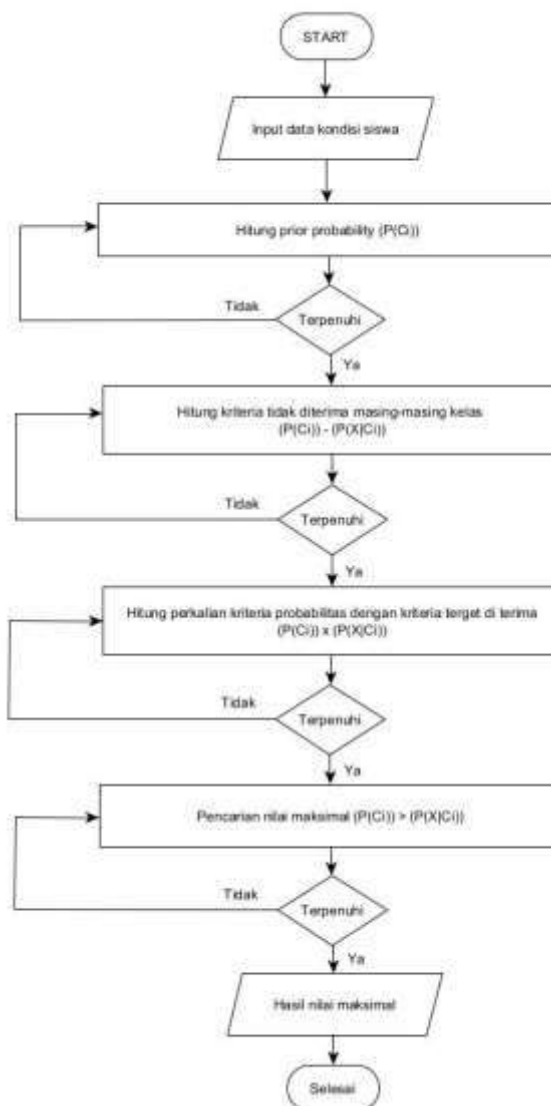
Setelah sudah diketahui hasil prediksi actual class, dalam tahap ini selanjutnya dilakukan pengamatan hasil prediksi, dimana hal-hal yang perlu diamati adalah nilai *precision*, nilai *recall* dan nilai *accuracy*.

1. Tahap yang terakhir adalah melakukan penulisan atau menyimpulkan hasil penelitian yang telah dilakukan. Beberapa hal yang akan disampaikan yaitu hasil prediksi dan nilai kinerja dari sistem prediksi yang telah dibangun.
2. Pada tahap ini dilakukan evaluasi dari model yang sudah diciptakan dari masing-masing algoritma.

BAB IV METODE NAIVE BAYES

4.1.Desain

Desain yaitu deskripsi atau definisi tentang langkah-langkah peneliti dalam melakukan pemecahan masalah yang akan diteliti. Melalui langkah-langkah pada desain penelitian, peneliti bisa menjelaskan secara jelas langkah-langkah apa saja yang diteliti dan dapat memberikan kesimpulan hasil. Berikut desain penelitian yang dibuat penulis bisa diperlihatkan pada Gambar 4.1.



Gambar 4. 1 Desain penelitian naïve bayes

Tentang Gambar 4.1 flowchart naïve bayes, menjelaskan proses yang dilakukan untuk menentukan nilai suatu kelas $h \in H$ dari suatu dokumen $x \in X$ dengan $H = \{h_1, h_2, h_3, \dots, h_p\}$ dan $X = \{x_1, x_2, x_3, \dots, x_q\}$. Sedangkan penentuan kelas dalam klasifikasi dokumen dilakukan dengan cara memiliki nilai maksimum dari $p(h|x)$ berdasarkan distribusi probabilitas.

$$P = \{p(h|x) | h \in H \text{ dan } x \in X\}$$

Sehingga dokumen x ke 1 dapat direpresentasikan sebagai vektor, dan nilai fitur dokumen yaitu $x = \{f_1, f_2, f_3, \dots, f_n\}$, nilai-nilai dari elemen setiap vector adalah nilai untuk fitur f_j pada himpunan fitur $F = \{f_1, f_2, f_3, \dots, f_n\}$ dengan f_{ij} adalah nilai dari fitur ke j pada dokumen x ke i .

Berdasarkan algoritma naïve bayes adalah persamaan perhitungan nilai dari probabilitas $p(h|x)$ sebagai berikut persamaannya (Arhami & Nasir, 2020).

$$Posterior = \frac{likelihood.Prior}{Evidance} \quad (4.1)$$

$$P(h|x) = \frac{P(x|h)*P(h)}{P(x)} \quad (4.2)$$

Keterangan :

$P(h|x)$ = Probabilitas hipotesis berdasarkan kondisi posterior

$P(h)$ = Probabilitas hipotesis prior probabilitas

$P(x|h)$ = Probabilitas berdasarkan kondisi h

$P(x)$ = Probabilitas x

Langkah selanjutnya dalam melakukan proses prediksi menggunakan metode *Naïve Bayes* diperlukan proses normalisasi data terlebih dahulu bertujuan untuk membersihkan data dari atribut yang kosong, menghapus data yang kosong, melakukan penghapusan data yang kurang penting dan hanya mengambil data siswa dan atribut dari data orang tua siswa yang berguna sebagai parameter prediksi pembayaran biaya pendidikan, yang dapat di lihat pada Tabel 3.12. Nama, kelamin, pekerjaan orang tua, penghasilan orang tua perbulan dan tanggungan orang tua.

Langkah selanjutnya setelah data ditentukan, menghitung jumlah tepat, terlambat dan tidak bayar dari tabel penelitian didapatkan berjumlah 420 dataset dengan 7 atribut dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat

sebanyak 72 data, dan tidak bayar sebanyak 48 data. Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*). Perhitungan probabilitas kemungkinan tepat waktu, terlambat dan tidak bayar menentukan faktor tunggakan pembayaran biaya pendidikan dilihat pada persamaan di bawah ini, yaitu : Yang terlambat membayar biaya pendidikan adalah :

$$P(\text{Terlambat}) = \frac{300}{315} = 0.9523$$

Sedangkan yang tepat waktu adalah :

$$P(\text{Tepat}) = \frac{72}{315} = 0.2285$$

dan yang tidak bayar adalah :

$$P(\text{Terlambat}) = \frac{48}{315} = 0.1523$$

Setelah menentukan probabilitas setiap prior, hitunglah probabilitas setiap kriteria dan komponen yang membentuk setiap kriteria. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan menghitung jumlah terlambat, tepat dan tidak bayar. sehingga perhitungan probabilitas setiap kriteria dapat dilihat pada tabel berikut.

Kemungkinan parameter pekerjaan orang tua , berdasarkan data yang dikumpulkan, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.1, tentang hasil probabilitas pekerjaan orang tua. Hasil perhitungan probabilitas pada parameter pekerjaan orang tua.

Tabel 4. 1 Pekerjaan Orang Tua

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak bayar
S	143	27	23	0.4766	0.375	0.4791
P	29	8	5	0.0966	0.1111	0.10417
T	38	11	7	0.1266	0.1527	0.1458
G	12	6	3	0.04	0.0833	0.0625
W	46	15	5	0.1533	0.2083	0.10417

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak bayar
PN	32	5	5	0.1066	0.0694	0.1041

Probabilitas parameter penghasilan orang tua per bulan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.2, tentang hasil probabilitas penghasilan orang tua per bulan. Hasil perhitungan probabilitas pada parameter penghasilan orang tua per bulan.

Tabel 4. 2 Penghasilan orang tua per bulan

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak bayar
C	55	17	10	0.1833	0.2361	0.2083
CT	127	23	21	0.4233	0.3194	0.4375
T	117	31	16	0.39	0.4305	0.3333
ST	1	1	1	0.0033	0.0138	0.0208

Probabilitas parameter jumlah tanggungan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.3, tentang hasil probabilitas jumlah tanggungan. Hasil perhitungan probabilitas pada parameter jumlah tanggungan.

Tabel 4. 3 Jumlah Tanggungan

Jumlah Tanggungan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak bayar
C	224	56	37	0.7466	0.7777	0.7708
B	76	16	11	0.25333	0.2222	0.2291

Langkah selanjutnya adalah menentukan probabilitas masing-masing kriteria, proses perhitungan data testing no. 316.

316	NABILA SANDRA AURELYA RAMADHANI	P	P	CT	4	B	?
-----	---------------------------------	---	---	----	---	---	---

$$\begin{aligned}
P(336|Terlambat) &= P(\text{Pekerjaan orang tua} \\
&= S|Terlambat) \times P(\text{Penghasilan orang tua} \\
&= T|Terlambat) \times P(\text{Jumlah tanggungan} = C|Terlambat) \\
&= 0.4766 \times 0.39 \times 0.7466 = 0.1387
\end{aligned}$$

$$\begin{aligned}
P(336|Tepat) &= P(\text{Pekerjaan orang tua} \\
&= S|Tepat) \times P(\text{Penghasilan orang tua} \\
&= T|Tepat) \times P(\text{Jumlah tanggungan} = C|Tepat) \\
&= 0.1111 \times 0.3194 \times 0.2222 = 0.0078
\end{aligned}$$

$$\begin{aligned}
P(336|Tidak bayar) &= P(\text{Pekerjaan orang tua} \\
&= S|Tidak bayar) \times P(\text{Penghasilan orang tua} \\
&= T|Tidak bayar) \times P(\text{Jumlah tanggungan} \\
&= C|Tidak bayar) = 0.1041 \times 0.4375 \times 0.2291 = 0.0104
\end{aligned}$$

Setelah nilai terlambat, tepat dan tidak bayar pada data 336 sampai dengan 420 diketahui. Selain itu, perhitungan maksimal untuk setiap klasifikasi dilakukan oleh penulis. Pengumpulan data dari responden 336 sampai dengan 420 untuk menentukan bagaimana memaksimalkan nilai probabilitas yaitu:

$$\begin{aligned}
\text{Hob Terlambat} &= \frac{P(336|Terlambat)}{P((336|Terlambat)P(336|Tepat) + P(336|Tidak bayar))} \\
&= \frac{0.1387}{0.1387 + 0.0078 + 0.0104} = 0.8840
\end{aligned}$$

$$\begin{aligned}
\text{Hob Tepat} &= \frac{P(336|Tepat)}{P((316|Terlambat) + P(316|Tepat) + P(316|Tidak bayar))} \\
&= \frac{0.0078}{0.1387 + 0.0078 + 0.0104} = 0.0497
\end{aligned}$$

Hob Tidak bayar

$$\begin{aligned}
 &= \frac{P(336|\textit{Tidak bayar})}{P((316|\textit{Terlambat}) + P(316|\textit{Tepat}) + P(316|\textit{Tidak bayar}))} \\
 &= \frac{0.0104}{0.1387 + 0.0078 + 0.0104} = 0.0662
 \end{aligned}$$

Hob Tidak bayar < Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = “**Terlambat**”.

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class tidak bayar yang memiliki nilai paling tinggi.

4.2.Uji Coba

Pada tahap selanjutnya yaitu melakukan pengujian penerapan algoritma naïve bayes dengan menerapkan 5 skenario percobaan. Percobaan pertama, melakukan perbandingan data uji 80:20 dengan semua parameter, percobaan kedua, melakukan perbandingan data uji 75:25 dengan semua parameter, percobaan ketiga, melakukan perbandingan data uji 70:30 dengan semua parameter, percobaan keempat, melakukan perbandingan data uji 60:40 dengan semua parameter, dan percobaan kelima, melakukan perbandingan data uji 50:50 dengan semua parameter. Dari penerapan kelima percobaan tersebut akan dihasilkan nilai performa algoritma naïve bayes (*accuracy*, *precision*, *recall* dan *f1-score* dari masing-masing percobaan. Berikut adalah langkah-langkah penerapan percobaan yang digunakan :

4.2.1. Percobaan Pertama Data Uji 80:20

Pada percobaan ini, data yang digunakan adalah data yang sama dengan Tabel 3.1.1. tentang tabel kondisi siswa, langkah selanjutnya menghitung jumlah tepat, terlambat dan tidak bayar berdasarkan tabel penelitian. Berikut percobaan pertama dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan 420 dataset, 3 atribut (pekerjaan orang tua, penghasilan orang tua per bulan, dan jumlah tanggungan) dan terbagi 3 class

yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada tabel 4.4:

Tabel 4. 4 Jumlah dataset uji coba 80:20 skenario pertama

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 *record*) sisanya akan dijadikan sebagai data *testing* 20% (84 *record*). Perhitungan probabilitas kemungkinan tepat waktu, terlambat dan tidak bayar menentukan faktor tunggakan pembayaran biaya pendidikan dilihat pada Tabel 4.5 :

Tabel 4. 5 Probabilitas actual class uji coba 80:20 skenario pertama

Probabilitas Actual Class	Terlambat	Tepat	Tidak Bayar
	$300/336 =$ 0.8928	$72/336 =$ 0.2142	$48/336 =$ 0.1428

Setelah probabilitas dari masing-masing prior dan probabilitas masing-masing actual class diketahui, Setelah itu, tentukan probabilitas masing-masing kriteria dan hitunglah komponen-komponen penyusun setiap kriteria yang menyusun kriteria. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan menghitung jumlah terlambat, tepat dan tidak bayar. Sehingga perhitungan probabilitas setiap kriteria dapat dilihat pada tabel berikut.

Probabilitas parameter pekerjaan orang tua , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.6, tentang hasil probabilitas pekerjaan orang tua. Hasil perhitungan probabilitas pada parameter pekerjaan orang tua.

Tabel 4. 6 Pekerjaan orang tua uji coba 80:20 skenario pertama

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak Bayar	Terlambat	Tepat	Tidak Bayar
S	143	27	23	0.4766	0.375	0.4791
P	29	8	5	0.0966	0.1111	0.10417

T	38	11	7	0.1266	0.1527	0.1458
G	12	6	3	0.04	0.0833	0.0625
W	46	15	5	0.1533	0.2083	0.10417
PN	32	5	5	0.1066	0.0694	0.1041

Probabilitas parameter penghasilan orang tua per bulan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.7, tentang hasil probabilitas penghasilan orang tua per bulan. Hasil perhitungan probabilitas pada parameter penghasilan orang tua per bulan.

Tabel 4. 7 Penghasilan orang tua per bulan uji coba 80:20 skenario pertama

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak Bayar
C	55	17	10	0.1833	0.2361	0.2083
CT	127	23	21	0.4233	0.3194	0.4375
T	117	31	16	0.39	0.4305	0.3333
ST	1	1	1	0.0033	0.0138	0.0208

Probabilitas parameter jumlah tanggungan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.8, tentang hasil probabilitas jumlah tanggungan. Hasil perhitungan probabilitas pada parameter jumlah tanggungan.

Tabel 4. 8 Jumlah Tanggungan uji coba 80:20 skenario pertama

Jumlah Tanggungan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak Bayar
C	224	56	37	0.7466	0.7777	0.7708
B	76	16	11	0.25333	0.2222	0.2291

Langkah selanjutnya adalah menentukan probabilitas masing-masing kriteria, proses perhitungan data testing no. 336.

336	NICO BARI' PRATAMA	L	S	T	C	?
-----	--------------------	---	---	---	---	---

$$P(336|Terlambat)$$

$$\begin{aligned} &= P(\text{Pekerjaan orang tua} \\ &= S|Terlambat) \times P(\text{Penghasilan orang tua} \\ &= T|Terlambat) \times P(\text{Jumlah tanggungan} = C|Terlambat) \\ &= 0.4766 \times 0.39 \times 0.7466 = 0.1387 \end{aligned}$$

$$P(336|Tepat) = P(\text{Pekerjaan orang tua}$$

$$\begin{aligned} &= S|Tepat) \times P(\text{Penghasilan orang tua} \\ &= T|Tepat) \times P(\text{Jumlah tanggungan} = C|Tepat) \\ &= 0.1111 \times 0.3194 \times 0.2222 = 0.0078 \end{aligned}$$

$$P(336|Tidak bayar)$$

$$\begin{aligned} &= P(\text{Pekerjaan orang tua} \\ &= S|Tidak bayar) \times P(\text{Penghasilan orang tua} \\ &= T|Tidak bayar) \times P(\text{Jumlah tanggungan} \\ &= C|Tidak bayar) = 0.1041 \times 0.4375 \times 0.2291 = 0.0104 \end{aligned}$$

Setelah nilai terlambat, tepat dan tidak bayar pada data 336 sampai dengan 420 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data responden 336 sampai dengan 420 untuk menghitung pemaksimalan nilai probabilitas yaitu:

Hob Terlambat

$$\begin{aligned} &= \frac{P(336|Terlambat)}{P((336|Terlambat) + P(336|Tepat) + P(336|Tidak bayar))} \\ &= \frac{0.1387}{0.1387 + 0.0078 + 0.0104} = 0.8840 \end{aligned}$$

$$\begin{aligned} \text{Hob Tepat} &= \frac{P(336|Tepat)}{P((316|Terlambat) + P(316|Tepat) + P(316|Tidak bayar))} \\ &= \frac{0.0078}{0.1387 + 0.0078 + 0.0104} = 0.0497 \end{aligned}$$

Hob Tidak bayar

$$= \frac{P(336|Tidak\ bayar)}{P((316|Terlambat) + P(316|Tepat) + P(316|Tidak\ bayar))}$$

$$= \frac{0.0104}{0.1387 + 0.0078 + 0.0104} = 0.0662$$

Hob Tidak bayar < Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = **“Terlambat”**.

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi.

Proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes dengan cara menentukan nilai akurasi, recall, presisi, dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah. Berikut tampilan dari *confusion matrix* ditunjukkan pada Tabel 4.9.

Tabel 4. 9 *Confusion matrix*

Correct Classification	Classification	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

- TP adalah jika hasil prediksi sama dengan nilai actual class
- TN adalah jika alternatif tidak di tunjuk oleh prediksi dan nilai actual class
- FP adalah jika alternatif disebutkan di nilai prediksi dan tidak disebutkan di nilai actual.
- FN adalah jika nilai actual tidak di tunjuk oleh hasil prediksi

Dari tabel 4.9, maka rumus untuk mencari akurasi, presisi, recall dan f1-score :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4.3)$$

$$Presisi = \frac{TP}{FN+TP} \times 100\% \quad (4.4)$$

$$Recall = \frac{TP}{FP+TP} \times 100\% \quad (4.5)$$

$$F1 - score = \frac{2x(Recall \times Precision)}{(Recall+Precision)} \times 100\% \quad (4.6)$$

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes yang ditampilkan pada hasil confusion matriks, ditunjukkan pada Tabel 4.10.

Tabel 4. 10 Hasil *confusion matrix* uji coba 80:20 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak Bayar
Terlambat	0	16	0
Tepat	0	57	0
Tidak Bayar	0	11	0
Total	0	84	0

Berdasarkan Tabel 4.11 tentang hasil *confusion matrix* percobaan 80:20 skenario pertama metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+57+0}{84} \times 100\% = 67\%$$

$$Presisi = \frac{57}{27+57} \times 100\% = 68\%$$

$$Recall = \frac{57}{0+57} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 68)}{1+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode naïve bayes diperoleh = 67%, presisi= 68%, dan recall = 100%, serta F1-score = 81%, dapat dilihat pada Tabel 4.11.

Tabel 4. 11 Hasil uji coba 80:20 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	67%
		2. Penghasilan	Tepat	0.68	1.00	0.81	
		3. Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

b. Skenario Kedua

Skenario kedua dilakukan dengan 420 dataset, 2 atribut (pekerjaan orang tua dan penghasilan orang tua per bulan) dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada Tabel 4.12:

Tabel 4. 12 Jumlah dataset uji coba 80:20 skenario kedua

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 record) sisanya akan dijadikan sebagai data *testing* 20% (84 record). Perhitungan probabilitas kemungkinan tepat waktu, terlambat dan tidak bayar menentukan faktor tunggakan pembayaran biaya pendidikan dilihat pada Tabel 4.13 :

Tabel 4. 13 Probabilitas actual class uji coba 80:20 skenario kedua

Probabilitas Actual Class	Terlambat	Tepat	Tidak Bayar
	$300/336 = 0.8928$	$72/336 = 0.2142$	$48/336 = 0.1428$

Setelah probabilitas dari masing-masing prior dan probabilitas masing-masing actual class diketahui. Setelah itu, tentukan probabilitas masing-masing kriteria dan hitunglah komponen-komponen penyusun setiap kriteria yang menyusun kriteria. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan

menghitung jumlah terlambat, tepat dan tidak bayar. Sehingga perhitungan probabilitas setiap kriteria dapat dilihat pada tabel berikut.

Probabilitas parameter pekerjaan orang tua , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.14, tentang hasil probabilitas pekerjaan orang tua. Hasil perhitungan probabilitas pada parameter pekerjaan orang tua.

Tabel 4. 14 Pekerjaan orang tua uji coba 80:20 skenario kedua

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak Bayar	Terlambat	Tepat	Tidak Bayar
S	143	27	23	0.4766	0.375	0.4791
P	29	8	5	0.0966	0.1111	0.10417
T	38	11	7	0.1266	0.1527	0.1458
G	12	6	3	0.04	0.0833	0.0625
W	46	15	5	0.1533	0.2083	0.10417
PN	32	5	5	0.1066	0.0694	0.1041

Probabilitas parameter penghasilan orang tua per bulan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.15, tentang hasil probabilitas penghasilan orang tua per bulan. Hasil perhitungan probabilitas pada parameter penghasilan orang tua per bulan.

Tabel 4. 15 Penghasilan orang tua per bulan uji coba 80:20 skenario kedua

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak Bayar	Terlambat	Tepat	Tidak Bayar
C	55	17	10	0.1833	0.2361	0.2083
CT	127	23	21	0.4233	0.3194	0.4375
T	117	31	16	0.39	0.4305	0.3333
ST	1	1	1	0.0033	0.0138	0.0208

Langkah selanjutnya adalah menentukan probabilitas masing-masing kriteria, proses perhitungan data testing no. 336.

336	NICO BARI PRATAMA	S	T	?
-----	-------------------	---	---	---

$$\begin{aligned}
P(336|Terlambat) &= P(\text{Pekerjaan orang tua} \\
&= S|Terlambat) \times P(\text{Penghasilan orang tua} = T|Terlambat) \\
&= 0.4766 \times 0.39 = 0.1858
\end{aligned}$$

$$\begin{aligned}
P(336|Tepat) &= P(\text{Pekerjaan orang tua} \\
&= S|Tepat) \times P(\text{Penghasilan orang tua} = T|Tepat) \\
&= 0.1111 \times 0.3194 = 0.0354
\end{aligned}$$

$$\begin{aligned}
P(336|Tidak bayar) &= P(\text{Pekerjaan orang tua} \\
&= S|Tidak bayar) \times P(\text{Penghasilan orang tua} = T|Tidak bayar) \\
&= 0.1041 \times 0.4375 = 0.0455
\end{aligned}$$

Setelah nilai terlambat, tepat dan tidak bayar pada data 336 sampai dengan 420 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data responden 336 sampai dengan 420 untuk menghitung pemaksimalan nilai probabilitas yaitu:

Hob Terlambat

$$\begin{aligned}
&= \frac{P(336|Terlambat)}{P((336|Terlambat) + P(336|Tepat) + P(336|Tidak bayar))} \\
&= \frac{0.1858}{0.18587 + 0.0354 + 0.0455} = 0.6966
\end{aligned}$$

$$\begin{aligned}
\text{Hob Tepat} &= \frac{P(336|Tepat)}{P((316|Terlambat) + P(316|Tepat) + P(316|Tidak bayar))} \\
&= \frac{0.0354}{0.1858 + 0.0354 + 0.0455} = 0.1327
\end{aligned}$$

Hob Tidak bayar

$$\begin{aligned}
&= \frac{P(336|Tidak bayar)}{P((316|Terlambat) + P(316|Tepat) + P(316|Tidak bayar))} \\
&= \frac{0.0455}{0.1858 + 0.0354 + 0.0455} = 0.1706
\end{aligned}$$

Hob Tidak bayar < Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = “**Terlambat**”.

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi.

Proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.16.

Tabel 4. 16 Hasil *confusion matrix* percobaan 80:20 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak Bayar
Terlambat	0	16	0
Tepat	0	57	0
Tidak Bayar	0	11	0
Total	0	84	0

Berdasarkan Tabel 4.16 tentang hasil *confusion matrix* percobaan 80:20 skenario kedua metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+57+0}{84} \times 100\% = 67\%$$

$$Presisi = \frac{57}{27+57} \times 100\% = 68\%$$

$$Recall = \frac{57}{0+57} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 68)}{1+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode naïve bayes diperoleh = 67%, presisi= 68%, dan recall = 100%, serta F1-score = 81%, dapat dilihat pada Tabel 4.17.

Tabel 4. 17 Hasil uji coba 80:20 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presi	Recal 1	F1-Score	
80%	20%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	67%
			Tepat	0.68	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

c. Skenario Ketiga

Skenario ketiga dilakukan dengan 420 dataset, 2 atribut (pekerjaan orang tua dan jumlah tanggungan) dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada Tabel 4.18:

Tabel 4. 18 Jumlah dataset uji coba 80:20 skenario ketiga

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 record) sisanya akan dijadikan sebagai data *testing* 20% (84 record). Perhitungan probabilitas kemungkinan tepat waktu, terlambat dan tidak bayar menentukan faktor tunggakan pembayaran biaya pendidikan dilihat pada Tabel 4.19 :

Tabel 4. 19 Probabilitas actual class uji coba 80:20 skenario ketiga

Probabilitas Actual Class	Terlambat	Tepat	Tidak Bayar
	$300/336 = 0.8928$	$72/336 = 0.2142$	$48/336 = 0.1428$

Setelah probabilitas dari masing-masing prior dan probabilitas masing-masing actual class telah diketahui. Setelah itu, tentukan probabilitas masing-masing kriteria dan hitunglah komponen-komponen penyusun setiap kriteria yang menyusun kriteria. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan menghitung jumlah terlambat, tepat dan tidak bayar. Sehingga perhitungan probabilitas setiap kriteria dapat dilihat pada tabel berikut.

Probabilitas parameter pekerjaan orang tua , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.20, tentang hasil probabilitas pekerjaan orang tua. Hasil perhitungan probabilitas pada parameter pekerjaan orang tua.

Tabel 4. 20 Pekerjaan orang tua uji coba 80:20 skenario ketiga

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak Bayar	Terlambat	Tepat	Tidak Bayar
S	143	27	23	0.4766	0.375	0.4791
P	29	8	5	0.0966	0.1111	0.10417
T	38	11	7	0.1266	0.1527	0.1458
G	12	6	3	0.04	0.0833	0.0625
W	46	15	5	0.1533	0.2083	0.10417
PN	32	5	5	0.1066	0.0694	0.1041

Probabilitas parameter jumlah tanggungan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.21, tentang hasil probabilitas jumlah tanggungan. Hasil perhitungan probabilitas pada parameter jumlah tanggungan.

Tabel 4. 21 Jumlah Tanggungan uji coba 80:20 skenario ketiga

Jumlah Tanggungan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak Bayar
C	224	56	37	0.7466	0.7777	0.7708
B	76	16	11	0.25333	0.2222	0.2291

Langkah selanjutnya adalah menentukan probabilitas masing-masing kriteria, proses perhitungan data testing no. 336.

336	NICO BARI' PRATAMA	S	C	?
-----	--------------------	---	---	---

$$P(336|Terlambat)$$

$$= P(\text{Pekerjaan orang tua}$$

$$= S|Terlambat) \times P(\text{Jumlah tanggungan} = C|Terlambat)$$

$$= 0.4766 \times 0.7466 = 0.3558$$

$$\begin{aligned}
 P(336|Tepat) &= P(\text{Pekerjaan orang tua} = S|Tepat) \times P(\text{Jumlah tanggungan} = C|Tepat) \\
 &= 0.1111 \times 0.2222 = 0.0246
 \end{aligned}$$

$$\begin{aligned}
 P(336|\text{Tidak bayar}) &= P(\text{Pekerjaan orang tua} = S|\text{Tidak bayar}) \times P(\text{Jumlah tanggungan} = C|\text{Tidak bayar}) \\
 &= 0.1041 \times 0.2291 = 0.0238
 \end{aligned}$$

Setelah nilai terlambat, tepat dan tidak bayar pada data 336 sampai dengan 420 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data responden 336 sampai dengan 420 untuk menghitung pemaksimalan nilai probabilitas yaitu:

Hob Terlambat

$$\begin{aligned}
 &= \frac{P(336|\text{Terlambat})}{P((336|\text{Terlambat}) + P(336|Tepat) + P(336|\text{Tidak bayar}))} \\
 &= \frac{0.3558}{0.3558 + 0.0246 + 0.0238} = 0.8802
 \end{aligned}$$

$$\begin{aligned}
 \text{Hob Tepat} &= \frac{P(336|Tepat)}{P((316|\text{Terlambat}) + P(316|Tepat) + P(316|\text{Tidak bayar}))} \\
 &= \frac{0.0246}{0.3558 + 0.0246 + 0.0238} = 0.0608
 \end{aligned}$$

Hob Tidak bayar

$$\begin{aligned}
 &= \frac{P(336|\text{Tidak bayar})}{P((316|\text{Terlambat}) + P(316|Tepat) + P(316|\text{Tidak bayar}))} \\
 &= \frac{0.0238}{0.3558 + 0.0246 + 0.0238} = 0.0588
 \end{aligned}$$

Hob Tidak bayar < Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = **“Terlambat”**.

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi.

Proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.22.

Tabel 4. 22 Hasil *confusion matrix* uji coba 80:20 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak Bayar
Terlambat	0	16	0
Tepat	0	57	0
Tidak Bayar	0	11	0
Total	0	84	0

Berdasarkan Tabel 4.22 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+57+0}{84} \times 100\% = 67\%$$

$$Presisi = \frac{57}{27+57} \times 100\% = 68\%$$

$$Recall = \frac{57}{0+57} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 68)}{1+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode naïve bayes diperoleh = 67%, presisi= 68%, dan recall = 100%, serta F1-score = 81%, dapat dilihat pada Tabel 4.23.

Tabel 4. 23 Hasil uji coba 80:20 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recal 1	F1-Score	
80%	20%	1.Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	67%
			Tepat	0.68	1.00	0.81	

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recal I	F1-Score	
		2.Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

d. Skenario Keempat

Skenario keempat dilakukan dengan 420 dataset, 3 atribut (penghasilan orang tua per bulan, dan jumlah tanggungan) dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada Tabel 4.24:

Tabel 4. 24 Jumlah dataset uji coba 80:20 skenario keempat

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 record) sisanya akan dijadikan sebagai data *testing* 20% (84 record). Perhitungan probabilitas kemungkinan tepat waktu, terlambat dan tidak bayar menentukan faktor tunggakan pembayaran biaya pendidikan dilihat pada Tabel 4.25 :

Tabel 4. 25 Probabilitas actual class uji coba 80:20 skenario keempat

Probabilitas Actual Class	Terlambat	Tepat	Tidak Bayar
	$300/336 = 0.8928$	$72/336 = 0.2142$	$48/336 = 0.1428$

Setelah probabilitas dari masing-masing prior dan probabilitas masing-masing actual class telah diketahui. Setelah itu, tentukan probabilitas masing-masing kriteria dan hitunglah komponen-komponen penyusun setiap kriteria yang menyusun kriteria. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan menghitung jumlah terlambat, tepat dan tidak bayar. Sehingga perhitungan probabilitas setiap kriteria dapat dilihat pada tabel berikut.

Probabilitas parameter penghasilan orang tua per bulan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.26, tentang hasil probabilitas penghasilan orang tua per bulan. Hasil perhitungan probabilitas pada parameter penghasilan orang tua per bulan.

Tabel 4. 26 Penghasilan orang tua per bulan uji coba 80:20 skenario keempat

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak Bayar
C	55	17	10	0.1833	0.2361	0.2083
CT	127	23	21	0.4233	0.3194	0.4375
T	117	31	16	0.39	0.4305	0.3333
ST	1	1	1	0.0033	0.0138	0.0208

Probabilitas parameter jumlah tanggungan , berdasarkan data yang terkumpul, hasil perhitungan probabilitasnya dapat di lihat pada Tabel 4.27, tentang hasil probabilitas jumlah tanggungan. Hasil perhitungan probabilitas pada parameter jumlah tanggungan.

Tabel 4. 27 Jumlah Tanggungan uji coba 80:20 skenario keempat

Jumlah Tanggungan	Jumlah Kejadian Di Pilih			Probabilitas		
	Terlambat	Tepat	Tidak bayar	Terlambat	Tepat	Tidak bayar
C	224	56	37	0.7466	0.7777	0.7708
B	76	16	11	0.25333	0.2222	0.2291

Langkah selanjutnya adalah menentukan probabilitas masing-masing kriteria, proses perhitungan data testing no. 336.

336	NICO BARI' PRATAMA	T	C	?
-----	--------------------	---	---	---

$$P(336|Terlambat)$$

$$= P(\text{Penghasilan orang tua}$$

$$= T|Terlambat) \times P(\text{Jumlah tanggungan} = C|Terlambat)$$

$$= 0.39 \times 0.7466 = 0.2911$$

$$\begin{aligned}
 P(336|Tepat) &= P(\text{Penghasilan orang tua} \\
 &= CT|Tepat) \times P(\text{Jumlah tanggungan} = B|Tepat) \\
 &= 0.3194 \times 0.2222 = 0.0709
 \end{aligned}$$

$$\begin{aligned}
 P(336|\text{Tidak bayar}) &= P(\text{Penghasilan orang tua} \\
 &= CT|\text{Tidak bayar}) \times P(\text{Jumlah tanggungan} = B|\text{Tidak bayar}) \\
 &= 0.4375 \times 0.2291 = 0.1002
 \end{aligned}$$

Setelah nilai terlambat, tepat dan tidak bayar pada data 336 sampai dengan 420 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data responden 336 sampai dengan 420 untuk menghitung pemaksimalan nilai probabilitas yaitu:

Hob Terlambat

$$\begin{aligned}
 &= \frac{P(336|\text{Terlambat})}{P((336|\text{Terlambat}) + P(336|Tepat) + P(336|\text{Tidak bayar}))} \\
 &= \frac{0.2911}{0.2911 + 0.0709 + 0.1002} = 0.6298
 \end{aligned}$$

$$\begin{aligned}
 \text{Hob Tepat} &= \frac{P(336|Tepat)}{P((316|\text{Terlambat}) + P(316|Tepat) + P(316|\text{Tidak bayar}))} \\
 &= \frac{0.0709}{0.2911 + 0.0709 + 0.1002} = 0.1533
 \end{aligned}$$

Hob Tidak bayar

$$\begin{aligned}
 &= \frac{P(336|\text{Tidak bayar})}{P((316|\text{Terlambat}) + P(316|Tepat) + P(316|\text{Tidak bayar}))} \\
 &= \frac{0.1002}{0.2911 + 0.0709 + 0.1002} = 0.2167
 \end{aligned}$$

Hob Tidak bayar < Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = **“Terlambat”**.

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi.

Proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.28.

Tabel 4. 28 Hasil *confusion matrix* uji coba 80:20 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak Bayar
Terlambat	0	16	0
Tepat	0	57	0
Tidak Bayar	0	11	0
Total	0	84	0

Berdasarkan Tabel 4.28 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+57+0}{84} \times 100\% = 67\%$$

$$Presisi = \frac{57}{27+57} \times 100\% = 68\%$$

$$Recall = \frac{57}{0+57} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 68)}{1+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode naïve bayes diperoleh = 67%, presisi= 68%, dan recall = 1%, serta F1-score = 81%, dapat dilihat pada Tabel 4.29.

Tabel 4. 29 Hasil uji coba 80:20 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Penghasilan 2. Jumlah	Terlambat	0.00	0.00	0.00	67%
			Tepat	0.68	1.00	0.81	

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
		tanggungan	Tidak Bayar	0.00	0.00	0.00	

4.2.2. Percobaan Kedua Data Uji 75:25

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan kedua dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 record) sisanya akan dijadikan sebagai data *testing* 25% (105 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengevaluasi kinerja dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.30.

Tabel 4. 30 Hasil *confusion matrix* uji coba 75:25 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	0	72	0
Tidak Bayar	0	13	0
Total	0	105	0

Berdasarkan Tabel 4.30 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+72+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{72}{0+33} \times 100\% = 69\%$$

$$Recall = \frac{72}{72+0} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 81)}{1+81} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode *naïve bayes* , presisi = 69%, recall= 100% , dan F1-score= 81%, dapat dilihat pada Tabel 4.31.

Tabel 4. 31 Hasil uji coba 75:25 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Traing	Testing			Presisi	Recall	F1-Score	
75%	25%	1. Pekerjaan orang tua 2. Penghasilan 3. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	68%
			Tepat	0.69	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai

performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.32.

Tabel 4. 32 Hasil *confusion matrix* uji coba 75:25 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	0	72	0
Tidak Bayar	0	13	0
Total	0	105	0

Berdasarkan Tabel 4.32 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+72+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{72}{33+0} \times 100\% = 69\%$$

$$Recall = \frac{72}{0+72} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 81)}{1+81} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode naïve bayes , presisi = 69%, recall= 1% , dan F1-score= 81%, dapat dilihat pada Tabel 4.33.

Tabel 4. 33 Hasil uji coba 75:25 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	68%
			Tepat	0.69	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

c. Skenario Ketiga

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.34.

Tabel 4. 34 Hasil *confusion matrix* uji coba 75:25 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	0	72	0
Tidak Bayar	0	13	0
Total	0	105	0

Berdasarkan Tabel 4.34 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+72+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{72}{0+33} \times 100\% = 69\%$$

$$Recall = \frac{72}{72+0} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 81)}{1+81} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode *naïve bayes* , presisi = 69%, recall= 1% , dan F1-score= 81%, dapat dilihat pada Tabel 4.35.

Tabel 4. 35 Hasil uji coba 75:25 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1.Pekerjaan orang tua 2.Jumlah tanggungan	Terlambat	0.00	0.00	0.00	68%
			Tepat	0.69	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode *naïve bayes* dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.36.

Tabel 4. 36 Hasil *confusion matrix* uji coba 75:25 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	0	72	0
Tidak Bayar	0	13	0
Total	0	105	0

Berdasarkan Tabel 4.36 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+72+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{72}{0+33} \times 100\% = 69\%$$

$$Recall = \frac{72}{72+0} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 81)}{1+81} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode *naive bayes* , presisi = 69%, recall= 100% , dan F1-score= 81%, dapat dilihat pada Tabel 4.37.

Tabel 4. 37 Hasil uji coba 75:25 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1.Penghasilan 2.Jumlah tanggungan	Terlambat	0.00	0.00	0.00	68%
			Tepat	0.69	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

4.2.3. Percobaan Ketiga Data Uji 70:30

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan kedua dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama. Pada skenario pertama dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan)

dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 *record*) sisanya akan dijadikan sebagai data *testing* 30% (126 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.38.

Tabel 4. 38 Hasil *confusion matrix* uji coba 70:30 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	25	0
Tepat	0	84	0
Tidak Bayar	0	17	0
Total	0	126	0

Berdasarkan Tabel 4.38 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+84+0}{126} \times 100\% = 67\%$$

$$Presisi = \frac{84}{42+84} \times 100\% = 67\%$$

$$Recall = \frac{84}{42+84} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 67)}{1+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode naïve bayes , presisi = 67%, recall= 1% , dan F1-score= 80%.,dapat dilihat pada Tabel 4.39.

Tabel 4. 39 Hasil uji coba 70:30 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	66%
		2. Penghasilan	Tepat	0.67	1.00	0.80	
		3. Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama. Pada skenario kedua dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan) dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 *record*) sisanya akan dijadikan sebagai data *testing* 30% (126 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.40.

Tabel 4. 40 Hasil *confusion matrix* uji coba 70:30 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	25	0
Tepat	0	84	0
Tidak Bayar	0	17	0
Total	0	126	0

Berdasarkan Tabel 4.31 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+84+0}{126} \times 100\% = 67\%$$

$$Presisi = \frac{84}{42+84} \times 100\% = 67\%$$

$$Recall = \frac{84}{42+84} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 67)}{1+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode *naïve bayes*, presisi = 67%, recall= 1%, dan F1-score= 80%, dapat dilihat pada Tabel 4.41.

Tabel 4. 41 Hasil uji coba 70:30 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.67	1.00	0.80	
			Tidak Bayar	0.00	0.00	0.00	

c. Skenario Ketiga

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama. Pada skenario ketiga dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 *record*) sisanya akan dijadikan sebagai data *testing* 30% (126 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode *naïve bayes* dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.42.

Tabel 4. 42 Hasil *confusion matrix* uji coba 70:30 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	25	0
Tepat	0	84	0
Tidak Bayar	0	17	0
Total	0	126	0

Berdasarkan Tabel 4.42 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+84+0}{126} \times 100\% = 67\%$$

$$Presisi = \frac{84}{42+84} \times 100\% = 67\%$$

$$Recall = \frac{84}{42+84} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 67)}{1+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode *naïve bayes* , presisi = 67%, recall= 1% , dan F1-score= 80%, dapat dilihat pada Tabel 4.43.

Tabel 4. 43 Hasil uji coba 70:30 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	66%
		2. Jumlah tanggungan	Tepat	0.67	1.00	0.80	
			Tidak Bayar	0.00	0.00	0.00	

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama. Pada skenario keempat dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 *record*) sisanya akan dijadikan sebagai data *testing* 30% (126 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.44.

Tabel 4. 44 Hasil *confusion matrix* uji coba 70:30 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	25	0
Tepat	0	84	0
Tidak Bayar	0	17	0
Total	0	126	0

Berdasarkan Tabel 4.44 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+84+0}{126} \times 100\% = 67\%$$

$$Presisi = \frac{84}{42+84} \times 100\% = 67\%$$

$$Recall = \frac{84}{42+84} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 67)}{1+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode *naïve bayes* , presisi = 67%, recall= 1% , dan F1-score= 80%, dapat dilihat pada Tabel 4.45.

Tabel 4. 45 Hasil uji coba 70:30 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1.Penghasilan 2.Jumlah tanggungan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.67	1.00	0.80	
			Tidak Bayar	0.00	0.00	0.00	

4.2.4. Percobaan Keempat Data Uji 60:40

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan keempat dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama. Pada skenario pertama dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode *naïve bayes* dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.46.

Tabel 4. 46 Hasil *confusion matrix* uji coba 60:40 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.46 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+117+0}{168} \times 100\% = 69\%$$

$$Presisi = \frac{117}{51+117} \times 100\% = 70\%$$

$$Recall = \frac{117}{0+117} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 70)}{1+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode *naive bayes*, presisi = 70%, recall= 100%, dan F1-score= 82%., dapat dilihat pada Tabel 4.47.

Tabel 4. 47 Hasil uji coba 60:40 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	69%
		2. Penghasilan	Tepat	0.70	1.00	0.82	
		3. Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama. Pada skenario kedua dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan) dataset dibagi menjadi 2, dengan

perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.48.

Tabel 4. 48 Hasil *confusion matrix* uji coba 60:40 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.48 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian:

$$Akurasi = \frac{0+117+0}{168} \times 100\% = 69\%$$

$$Presisi = \frac{117}{51+117} \times 100\% = 70\%$$

$$Recall = \frac{117}{0+117} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 70)}{1+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode naïve bayes , presisi = 70%, recall= 1% , dan F1-score= 82%, dapat dilihat pada Tabel 4.49.

Tabel 4. 49 Hasil uji coba 60:40 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	69%
			Tepat	0.70	1.00	0.82	
			Tidak Bayar	0.00	0.00	0.00	

c. Skenario Ketiga

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama. Pada skenario ketiga dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.50.

Tabel 4. 50 Hasil *confusion matrix* uji coba 60:40 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.50 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+117+0}{168} \times 100\% = 69\%$$

$$Presisi = \frac{117}{51+117} \times 100\% = 70\%$$

$$Recall = \frac{117}{0+117} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 70)}{1+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode *naïve bayes* , presisi = 70%, recall= 100% , dan F1-score= 82%, dapat dilihat pada Tabel 4.51.

Tabel 4. 51 Hasil uji coba 60:40 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	69%
		2. Jumlah tanggungan	Tepat	0.70	1.00	0.82	
			Tidak Bayar	0.00	0.00	0.00	

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama. Pada skenario keempat dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode *naïve bayes* dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.52.

Tabel 4. 52 Hasil *confusion matrix* uji coba 60:40 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.52 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+117+0}{168} \times 100\% = 69\%$$

$$Presisi = \frac{117}{51+117} \times 100\% = 70\%$$

$$Recall = \frac{117}{0+117} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 70)}{1+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode *naive bayes*, presisi = 70%, recall= 100%, dan F1-score= 82%, dapat dilihat pada Tabel 4.53.

Tabel 4. 53 Hasil uji coba 60:40 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	69%
			Tepat	0.70	1.00	0.82	
			Tidak Bayar	0.00	0.00	0.00	

4.2.5. Percobaan Kelima Data Uji 50:50

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan

percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan keempat dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama. Pada skenario pertama dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 50:50, data *training* 50% (210 *record*) sisanya akan dijadikan sebagai data *testing* 50% (210 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.54.

Tabel 4. 54 Hasil *confusion matrix* uji coba 50:50 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	34	0
Tepat	0	151	0
Tidak Bayar	0	25	0
Total	0	210	0

Berdasarkan Tabel 4.54 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian:

$$Akurasi = \frac{0+151+0}{210} \times 100\% = 71\%$$

$$Presisi = \frac{115}{59+115} \times 100\% = 72\%$$

$$Recall = \frac{151}{0+151} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 72)}{1+72} = 84\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 71 % dengan metode *naïve bayes*, presisi = 72%, recall= 1%, dan F1-score= 84%, dapat dilihat pada Tabel 4.55.

Tabel 4. 55 Hasil uji coba 50:50 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	71%
		2. Penghasilan	Tepat	0.72	1.00	0.84	
		3. Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama. Pada skenario kedua dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan) dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi *naïve bayes* dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode *naïve bayes* dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi *naïve bayes* yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.56.

Tabel 4. 56 Hasil *confusion matrix* uji coba 50:50 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.56 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+151+0}{210} \times 100\% = 71\%$$

$$Presisi = \frac{115}{59+115} \times 100\% = 72\%$$

$$Recall = \frac{151}{0+151} \times 100\% = 1\%$$

$$F1 - score = \frac{2 \times (1 \times 72)}{1+72} = 84\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 71 % dengan metode *naïve bayes*, presisi = 72%, recall= 1%, dan F1-score= 84%., dapat dilihat pada Tabel 4.57.

Tabel 4. 57 Hasil uji coba 50:50 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	71%
			Tepat	0.72	1.00	0.84	
			Tidak Bayar	0.00	0.00	0.00	

c. Skenario Ketiga

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama. Pada skenario ketiga dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan) dataset dibagi menjadi 2,

dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.58.

Tabel 4. 58 Hasil *confusion matrix* uji coba 50:50 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.58 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+151+0}{210} \times 100\% = 71\%$$

$$Presisi = \frac{115}{59+115} \times 100\% = 72\%$$

$$Recall = \frac{151}{0+151} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 72)}{1+72} = 84\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 71 % dengan metode naïve bayes , presisi = 72%, recall= 100% , dan F1-score= 84%, dapat dilihat pada Tabel 4.59.

Tabel 4. 59 Hasil uji coba 50:50 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	71%
		2. Jumlah tanggungan	Tepat	0.72	1.00	0.84	
			Tidak Bayar	0.00	0.00	0.00	

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama. Pada skenario keempat dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan) dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Langkah selanjutnya untuk mengetahui performa dari metode klasifikasi naïve bayes dengan cara menghitung nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode naïve bayes dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi naïve bayes yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 4.60.

Tabel 4. 60 Hasil *confusion matrix* uji coba 50:50 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 4.60 tentang hasil *confusion matrix* metode naïve bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian

$$Akurasi = \frac{0+151+0}{210} \times 100\% = 71\%$$

$$Presisi = \frac{115}{59+115} \times 100\% = 72\%$$

$$Recall = \frac{151}{0+151} \times 100\% = 100\%$$

$$F1 - score = \frac{2 \times (1 \times 72)}{1+72} = 84\%$$

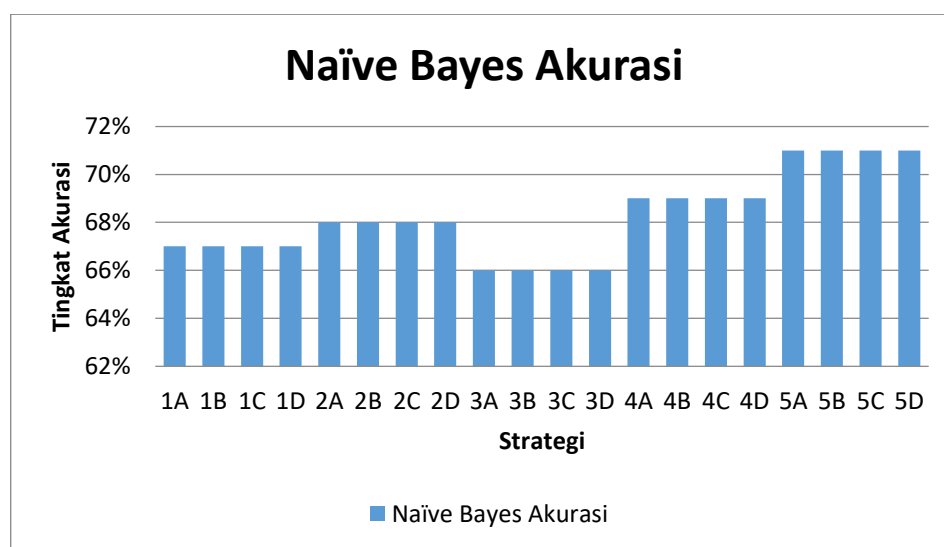
Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 71 % dengan metode *naïve bayes*, presisi = 72%, recall= 100%, dan F1-score= 84%, dapat dilihat pada Tabel 4.61.

Tabel 4. 61 Hasil uji coba 50:50 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	71%
			Tepat	0.72	1.00	0.84	
			Tidak Bayar	0.00	0.00	0.00	

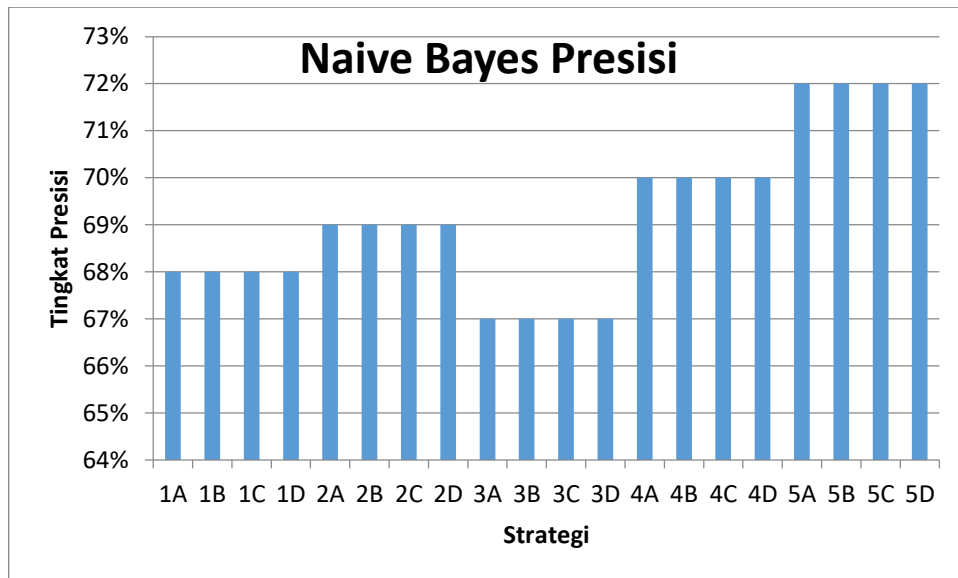
4.3. Kesimpulan

Setelah melakukan uji coba sebanyak 5 percobaan prediksi, dengan masing-masing percobaan terdapat 4 skenario menggunakan metode *naïve bayes*, maka dapat kita simpulkan pada Grafik 4.1.



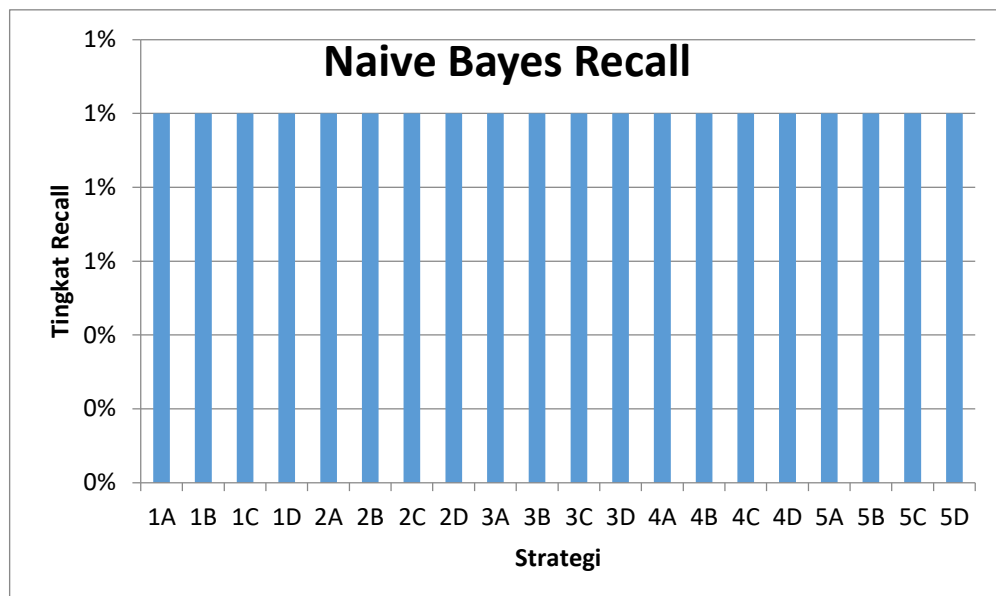
Grafik 4. 1 Hasil uji coba akurasi untuk semua skenario

Dari Grafik 4.1, di jelaskan bahwasannya akurasi tertinggi terdapat pada skenario 5A, 5B, 5C, dan 5D dengan nilai akurasi :71%.



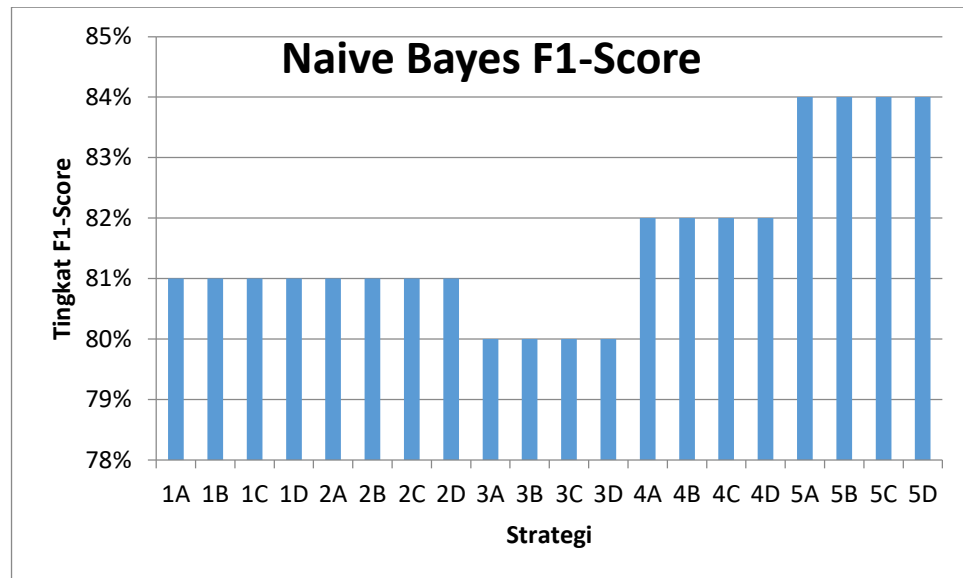
Grafik 4. 2 Hasil uji coba presisi untuk semua skenario

Dari hasil Grafik 4.2, dijelaskan bahwasannya hasil presisi tertinggi terdapat pada skenario 5A, 5B, 5C, dan 5D dengan nilai presisi: 72%.



Grafik 4. 3 Hasil uji coba recall untuk semua skenario

Dari hasil Grafik 4.3, dijelaskan bahwasannya hasil recall tertinggi terdapat pada semua skenario dengan nilai recall: 1%.



Grafik 4. 4 Hasil uji coba f1-score untuk semua skenario

Dari hasil Grafik 4.4, dijelaskan bahwasannya hasil f1-score tertinggi terdapat pada skenario 5A, 5B, 5C, dan 5D dengan nilai f1-score: 84%..

Dari grafik 4.1 sampai 4.4 yang ditampilkan, maka dapat di simpulkan skenario 5A, 5B, 5C, dan 5D menghasilkan akurasi tertinggi yaitu 71%, presisi = 72%, recall= 1% , dan F1-score= 84%, dengan parameter skenario 5A menggunakan semua parameter (pekerjaan orang tua, penghasilan orang tua, dan jumlah tanggungan orang tua), sedangkan pada skenario 5B menggunakan 2 parameter (pekerjaan orang tua dan penghasilan orang tua), skenario 5C menggunakan 2 parameter (pekerjaan orang tua dan jumlah tanggungan orang tua), dan skenario 5D menggunakan 2 parameter (penghasilan orang tua dan jumlah tanggungan orang tua).

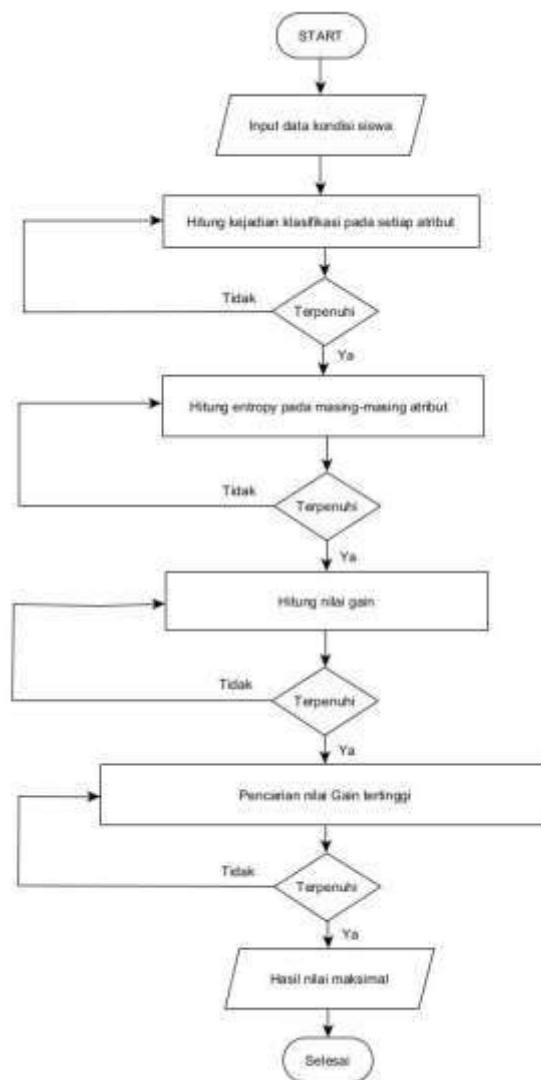
Jadi kesimpulannya pada penggunaan metode naïve bayes, dengan melakukan perbandingan data uji 50:50 dan menggunakan semua parameter, memiliki pengaruh dalam mengklasifikasi ekonomi siswa.

BAB V

METODE *DECISION TREE*

5.1.Desain

Desain yaitu deskripsi atau definisi tentang langkah-langkah peneliti dalam melakukan pemecahan masalah yang akan diteliti. Melalui langkah-langkah pada desain penelitian, peneliti bisa menjelaskan secara jelas langkah-langkah apa saja yang diteliti dan dapat memberikan kesimpulan hasil. Berikut desain penelitian yang dibuat penulis bisa dilihat pada Gambar 5.1.



Gambar 5. 1 Desain penelitian decision tree

Gambar 5.1. tentang flowchart decision tree, menjelaskan proses yang dilakukan untuk menentukan nilai suatu kelas. Sebuah obyek yang diklasifikasikan dalam pohon harus dites nilai entropy nya. Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity dan homogeneity dari kumpulan data. Dari nilai Entropy tersebut kemudian dihitung nilai information gain (IG) masing-masing atribut. Semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidakaslian S.sistem informasi atau disebut dengan processing system. Pemilihan atribut sebagai simpul, baik akar maupun simpul internal didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Rumus menghitung nilai entropy menggunakan persamaan (Arhami & Nasir, 2020):

$$Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i \quad (5.1)$$

Keterangan dari Persamaan :

S : himpunan kasus.

n : jumlah partisi S

Pi : proporsi Si terhadap S

Langkah selanjutnya yaitu menghitung gain. Information gain adalah salah satu attribute selection measure yang digunakan untuk memilih test attribute tiap node pada tree. Atribut dengan informasi gain tertinggi dipilih sebagai test atribut dari suatu node.

Gain (S,A) merupakan perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variable dependent S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (S,A). Menghitung nilai information gain menggunakan Persamaan :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (5.2)$$

Keterangan dari Persamaan :

S : himpunan kasus.

A : Atribut.

n : jumlah partisi atribut A.

|Si| : jumlah kasus pada partisi ke i.

|S| : jumlah kasus dalam S.

Langkah selanjutnya dalam melakukan proses prediksi menggunakan metode *decision tree* diperlukan proses normalisasi data terlebih dahulu bertujuan untuk membersihkan data dari atribut yang kosong, menghapus data yang kosong, melakukan penghapusan data yang kurang penting dan hanya mengambil data siswa dan atribut dari data orang tua siswa yang berguna sebagai parameter prediksi pembayaran biaya pendidikan, yang dapat di lihat pada Tabel 3.12. Nama, kelamin, pekerjaan orang tua, penghasilan orang tua perbulan dan tanggungan orang tua.

Langkah selanjutnya setelah data ditentukan, menghitung jumlah tepat, terlambat dan tidak bayar berdasarkan tabel penelitian didapatkan berjumlah 420 dataset dengan 7 atribut dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Data kondisi siswa tersebut lalu diklasifikasikan menjadi 3 bagian, yaitu pembagian berdasarkan pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan.

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian S (Terlambat) = 143, S (Tepat) = 27, S (Tidak bayar) = 23, hasil jumlah kejadian P (Terlambat) = 29, P (Tepat) = 8, P (Tidak bayar) = 5, hasil jumlah kejadian T (Terlambat) = 38, T (Tepat) = 11, T (Tidak bayar) = 7, hasil jumlah kejadian G (Terlambat) = 12, G (Tepat) = 6, G (Tidak bayar) = 3, hasil jumlah kejadian W (Terlambat) = 46, W (Tepat) = 15, W (Tidak bayar) = 5, sedangkan hasil jumlah kejadian PN (Terlambat) = 32, PN (Tepat) = 5, PN (Tidak bayar) = 5. Hasil klasifikasi dapat di lihat pada Tabel 5.1.

Tabel 5. 1 Klasifikasi berdasarkan pekerjaan orang tua

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
S	143	27	23
P	29	8	5
T	38	11	7
G	12	6	3
W	46	15	5
PN	32	5	5

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 55, C (Tepat) = 17, C (Tidak bayar) = 10, hasil jumlah kejadian CT (Terlambat) = 127, CT (Tepat) = 23, CT (Tidak bayar) = 21, hasil jumlah kejadian T (Terlambat) = 117, T (Tepat) = 31, T (Tidak bayar) = 16, sedangkan hasil jumlah kejadian ST (Terlambat) = 1, ST (Tepat) = 1, ST (Tidak bayar) = 1. Hasil klasifikasi dapat di lihat pada Tabel 5.2.

Tabel 5. 2 Klasifikasi berdasarkan penghasilan orang tua per bulan

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	55	17	10
CT	127	23	21
T	117	31	16
ST	1	1	1

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 224, C (Tepat) = 56, C (Tidak bayar) = 37, sedangkan hasil jumlah kejadian B (Terlambat) = 76, B (Tepat) = 16, B (Tidak bayar) = 11. Hasil klasifikasi dapat di lihat pada Tabel 5.3.

Tabel 5. 3 Klasifikasi berdasarkan jumlah tanggungan

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	224	56	37
B	76	16	11

Setelah masing-masing klasifikasi kejadian parameter telah diketahui, langkah selanjutnya adalah proses perhitungan entropy dan gain. Hasil entropy dan gain dapat di lihat pada Tabel 5.4.

Tabel 5. 4 Node awal

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain
Total		420	300	72	48	1.1405	
Pekerjaan Orang Tua	S	193	143	27	23	1.0832	0.0120
	P	42	29	8	5	1.1901	
	T	56	38	11	7	1.2158	
	G	21	12	6	3	1.3787	
	W	66	46	15	5	1.1308	
	PN	42	32	5	5	1.0299	
Penghasilan Orang Tua Per Bulan	C	82	55	17	10	1.2272	-0.4887
	CT	171	127	23	21	1.0795	
	T	164	117	31	16	1.1294	
	ST	3	1	1	1	1.5849	
Jumlah Tanggungan	C	317	224	56	37	1.1575	0.0006
	B	103	76	16	11	1.0855	

Pada perhitungan tabel node awal dapat diketahui bahwa atribut dengan Gain tertinggi adalah pekerjaan orang tua dengan nilai 0.0120, sehingga atribut pekerjaan orang tua dapat menjadi node akar.

5.2.Uji Coba

Pada tahap selanjutnya yaitu melakukan uji coba penerapan algoritma decision tree dengan menerapkan 5 skenario percobaan. Percobaan pertama, melakukan perbandingan data uji 80:20 dengan semua parameter, percobaan kedua, melakukan perbandingan data uji 75:25 dengan semua parameter, percobaan ketiga, melakukan perbandingan data uji 70:30 dengan semua parameter, percobaan keempat, melakukan perbandingan data uji 60:40 dengan

semua parameter, dan percobaan kelima, melakukan perbandingan data uji 50:50 dengan semua parameter. Dari penerapan kelima percobaan tersebut akan dihasilkan nilai performa algoritma naïve bayes (*accuracy*, *precision*, *recall* dan *f1-score* dari masing-masing percobaan. Berikut adalah langkah-langkah penerapan percobaan yang digunakan :

5.2.1. Percobaan Pertama Data Uji 80:20

Pada percobaan ini, data yang digunakan adalah data yang sama dengan Tabel 3.1.1. tentang tabel kondisi siswa, langkah selanjutnya menghitung jumlah tepat, terlambat dan tidak bayar berdasarkan tabel penelitian. Berikut percobaan pertama dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan 420 dataset, 3 atribut (pekerjaan orang tua, penghasilan orang tua per bulan, dan jumlah tanggungan) dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada tabel 5.5:

Tabel 5. 5 Jumlah dataset uji coba 80:20 skenario pertama

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 *record*) sisanya akan dijadikan sebagai data *testing* 20% (84 *record*). Data kondisi siswa tersebut lalu diklasifikasikan menjadi 4 bagian, yaitu pembagian berdasarkan pekerjaan orang tua, penghasilan orang tua per bulan, jumlah tanggungan.

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian S (Terlambat) = 143, S (Tepat) = 27, S (Tidak bayar) = 23, hasil jumlah kejadian P (Terlambat) = 29, P (Tepat) = 8, P (Tidak bayar) = 5, hasil jumlah kejadian T (Terlambat) = 38, T (Tepat) = 11, T (Tidak bayar) = 7, hasil jumlah kejadian G (Terlambat) = 12, G (Tepat) = 6, G (Tidak bayar) = 3, hasil jumlah kejadian W (Terlambat) = 46, W (Tepat) = 15, W (Tidak

bayar) = 5, sedangkan hasil jumlah kejadian PN (Terlambat) = 32, PN (Tepat) = 5, PN (Tidak bayar) = 5. Hasil klasifikasi dapat di lihat pada Tabel 5.6.

Tabel 5. 6 Klasifikasi pekerjaan orang tua uji coba 80:20 skenario pertama

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
S	143	27	23
P	29	8	5
T	38	11	7
G	12	6	3
W	46	15	5
PN	32	5	5

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 55, C (Tepat) = 17, C (Tidak bayar) = 10, hasil jumlah kejadian CT (Terlambat) = 127, CT (Tepat) = 23, CT (Tidak bayar) = 21, hasil jumlah kejadian T (Terlambat) = 117, T (Tepat) = 31, T (Tidak bayar) = 16, sedangkan hasil jumlah kejadian ST (Terlambat) = 1, ST (Tepat) = 1, ST (Tidak bayar) = 1. Hasil klasifikasi dapat di lihat pada Tabel 5.7.

Tabel 5. 7 Klasifikasi penghasilan orang tua uji coba 80:20 skenario pertama

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	55	17	10
CT	127	23	21
T	117	31	16
ST	1	1	1

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 224, C (Tepat) = 56, C (Tidak bayar) = 37, sedangkan hasil jumlah kejadian B (Terlambat) = 76, B (Tepat) = 16, B (Tidak bayar) = 11. Hasil klasifikasi dapat di lihat pada Tabel 5.8.

Tabel 5. 8 Klasifikasi jumlah tanggungan uji coba 80:20 skenario pertama

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	224	56	37

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
B	76	16	11

Setelah masing-masing klasifikasi kejadian parameter telah diketahui, langkah selanjutnya adalah proses perhitungan entropy dan gain. Hasil entropy dan gain dapat di lihat pada Tabel 5.9.

Tabel 5. 9 Node awal uji coba 80:20 skenario pertama

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain
Total		420	300	72	48	1.1405	
Pekerjaan Orang Tua	S	193	143	27	23	1.0832	0.0120
	P	42	29	8	5	1.1901	
	T	56	38	11	7	1.2158	
	G	21	12	6	3	1.3787	
	W	66	46	15	5	1.1308	
	PN	42	32	5	5	1.0299	
Penghasilan Orang Tua Per Bulan	C	82	55	17	10	1.2272	-0.4887
	CT	171	127	23	21	1.0795	
	T	164	117	31	16	1.1294	
	ST	3	1	1	1	1.5849	
Jumlah Tanggungan	C	317	224	56	37	1.1575	0.0006
	B	103	76	16	11	1.0855	

Pada perhitungan tabel node awal dapat diketahui bahwa atribut dengan Gain tertinggi adalah pekerjaan orang tua dengan nilai 0.0120, sehingga atribut pekerjaan orang tua dapat menjadi node akar.

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai

performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.10.

Tabel 5. 10 *Confusion matrix*

Correct Classification	Classification	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan :

- TP adalah jika hasil prediksi sama dengan nilai actual class
- TN adalah jika alternatif tidak di tunjuk oleh prediksi dan nilai actual class
- FP adalah jika alternatif disebutkan di nilai prediksi dan tidak disebutkan di nilai actual.
- FN adalah jika nilai actual tidak di tunjuk oleh hasil prediksi

Dari hasil proses klasifikasi dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.11.

Tabel 5. 11 Hasil *confusion matrix* uji coba 80:20 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	16	0
Tepat	1	56	0
Tidak Bayar	1	10	0
Total	2	82	0

Berdasarkan Tabel 5.11 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+56+0}{84} \times 100\% = 67\%$$

$$Recall = \frac{56}{1+56} \times 100\% = 98\%$$

$$Presisi = \frac{56}{26+56} \times 100\% = 68\%$$

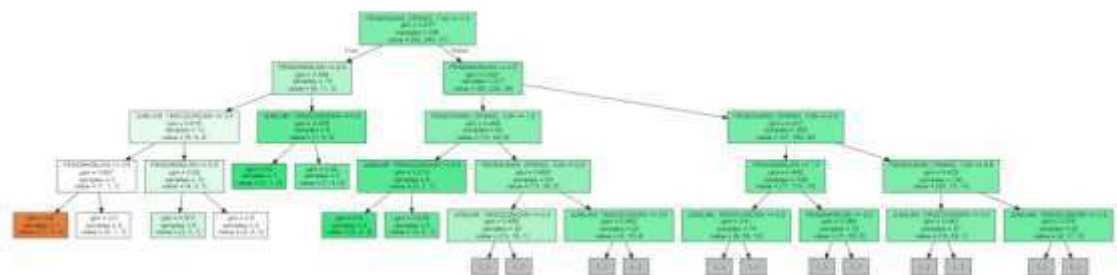
$$F1 - score = \frac{2 \times (98 \times 68)}{98+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode decision tree diperoleh = 67%, presisi= 68%, dan recall = 98%, serta F1-score = 81%, dapat dilihat pada Tabel 5.12.

Tabel 5. 12 Hasil uji coba 80:20 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Pekerjaan orang tua 2. Penghasilan 3. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.68	0.98	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.2.



Gambar 5. 2 Node akar percobaan 80:20 skenario pertama

b. Skenario Kedua

Skenario kedua dilakukan dengan 420 dataset, 2 atribut (pekerjaan orang tua dan penghasilan orang tua per bulan). dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada tabel 5.13:

Tabel 5. 13 Jumlah dataset uji coba 80:20 skenario kedua

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 *record*) sisanya akan dijadikan sebagai data *testing* 20% (84 *record*). Data kondisi siwa tersebut lalu diklasifikasikan menjadi 2 bagian, yaitu pembagian berdasarkan pekerjaan orang tua dan penghasilan orang tua per bulan.

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian S (Terlambat) = 143, S (Tepat) = 27, S (Tidak bayar) = 23, hasil jumlah kejadian P (Terlambat) = 29, P (Tepat) = 8, P (Tidak bayar) = 5, hasil jumlah kejadian T (Terlambat) = 38, T (Tepat) = 11, T (Tidak bayar) = 7, hasil jumlah kejadian G (Terlambat) = 12, G (Tepat) = 6, G (Tidak bayar) = 3, hasil jumlah kejadian W (Terlambat) = 46, W (Tepat) = 15, W (Tidak bayar) = 5, sedangkan hasil jumlah kejadian PN (Terlambat) = 32, PN (Tepat) = 5, PN (Tidak bayar) = 5. Hasil klasifikasi dapat di lihat pada Tabel 5.14.

Tabel 5. 14 Klasifikasi pekerjaan orang tua uji coba 80:20 skenario kedua

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
S	143	27	23
P	29	8	5
T	38	11	7
G	12	6	3
W	46	15	5
PN	32	5	5

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 55, C (Tepat) = 17, C (Tidak bayar) = 10, hasil jumlah kejadian CT (Terlambat) = 127, CT (Tepat) = 23, CT (Tidak bayar) = 21, hasil jumlah kejadian T (Terlambat) = 117, T (Tepat) = 31, T (Tidak bayar) = 16, sedangkan hasil jumlah kejadian ST (Terlambat) = 1, ST (Tepat) = 1, ST (Tidak bayar) = 1. Hasil klasifikasi dapat di lihat pada Tabel 5.15.

Tabel 5. 15 Klasifikasi penghasilan orang tua uji coba 80:20 skenario kedua

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	55	17	10
CT	127	23	21
T	117	31	16
ST	1	1	1

Setelah masing-masing klasifikasi kejadian parameter telah diketahui, langkah selanjutnya adalah proses perhitungan entropy dan gain. Hasil entropy dan gain dapat di lihat pada Tabel 5.16.

Tabel 5. 16 Node awal uji coba 80:20 skenario kedua

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain
Total		420	300	72	48	1.1405	
Pekerjaan Orang Tua	S	193	143	27	23	1.0832	0.0120
	P	42	29	8	5	1.1901	
	T	56	38	11	7	1.2158	
	G	21	12	6	3	1.3787	
	W	66	46	15	5	1.1308	
	PN	42	32	5	5	1.0299	
Penghasilan Orang Tua Per Bulan	C	82	55	17	10	1.2272	-0.4887
	CT	171	127	23	21	1.0795	
	T	164	117	31	16	1.1294	
	ST	3	1	1	1	1.5849	

Pada perhitungan tabel node awal dapat diketahui bahwa atribut dengan Gain tertinggi adalah pekerjaan orang tua dengan nilai 0.0120, sehingga atribut pekerjaan orang tua dapat menjadi node akar.

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.17.

Tabel 5. 17 Hasil *confusion matrix* uji coba 80:20 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	16	0
Tepat	1	56	0
Tidak Bayar	1	10	0
Total	2	82	0

Berdasarkan Tabel 5.17 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+56+0}{84} \times 100\% = 67\%$$

$$Recall = \frac{56}{1+56} \times 100\% = 98\%$$

$$Presisi = \frac{56}{26+56} \times 100\% = 68\%$$

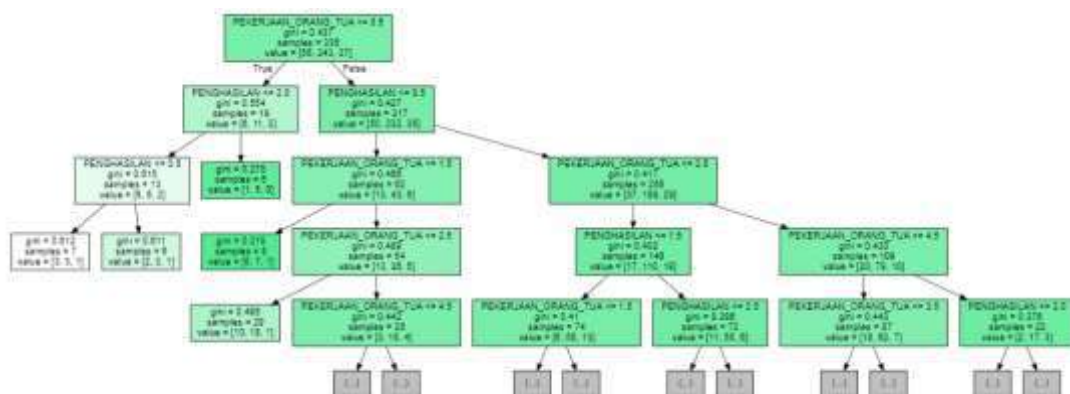
$$F1 - score = \frac{2 \times (98 \times 68)}{98+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode decision tree diperoleh = 67%, presisi= 68%, dan recall = 98%, serta F1-score = 81%, dapat di lihat pada Tabel 5.18.

Tabel 5. 18 Hasil uji coba 80:20 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.68	0.98	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.3.



Gambar 5. 3 Node akar percobaan 80:20 skenario kedua

c. Skenario Ketiga

Skenario ketiga dilakukan dengan 420 dataset, 2 atribut (pekerjaan orang tua dan jumlah tanggungan). dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada tabel 5.19:

Tabel 5. 19 Jumlah dataset uji coba 80:20 skenario ketiga

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 *record*) sisanya akan dijadikan sebagai data *testing* 20% (84 *record*). Data kondisi siwa tersebut lalu diklasifikasikan menjadi 2 bagian, yaitu pembagian berdasarkan pekerjaan orang tua dan jumlah tanggungan.

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian S (Terlambat) = 143, S (Tepat) = 27, S (Tidak bayar) = 23, hasil jumlah kejadian P (Terlambat) = 29, P (Tepat) = 8, P (Tidak bayar) = 5, hasil jumlah kejadian T (Terlambat) = 38, T (Tepat) = 11, T (Tidak bayar) = 7, hasil jumlah kejadian G (Terlambat) = 12, G (Tepat) = 6, G (Tidak bayar) = 3, hasil jumlah kejadian W (Terlambat) = 46, W (Tepat) = 15, W (Tidak bayar) = 5, sedangkan hasil jumlah kejadian PN (Terlambat) = 32, PN (Tepat) = 5, PN (Tidak bayar) = 5. Hasil klasifikasi dapat di lihat pada Tabel 5.20.

Tabel 5. 20 Klasifikasi pekerjaan orang tua uji coba 80:20 skenario ketiga

Pekerjaan Orang Tua	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
S	143	27	23
P	29	8	5
T	38	11	7
G	12	6	3
W	46	15	5
PN	32	5	5

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 224, C (Tepat) = 56, C (Tidak bayar) = 37, sedangkan hasil jumlah kejadian B (Terlambat) = 76, B (Tepat) = 16, B (Tidak bayar) = 11. Hasil klasifikasi dapat di lihat pada Tabel 5.21.

Tabel 5. 21 Klasifikasi jumlah tanggungan uji coba 80:20 skenario ketiga

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	224	56	37
B	76	16	11

Setelah masing-masing klasifikasi kejadian parameter telah diketahui, langkah selanjutnya adalah proses perhitungan entropy dan gain. Hasil entropy dan gain dapat di lihat pada Tabel 5.22.

Tabel 5. 22 Node awal uji coba 80:20 skenario ketiga

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain
Total		420	300	72	48	1.1405	
Pekerjaan Orang Tua	S	193	143	27	23	1.0832	0.0120
	P	42	29	8	5	1.1901	
	T	56	38	11	7	1.2158	
	G	21	12	6	3	1.3787	
	W	66	46	15	5	1.1308	
	PN	42	32	5	5	1.0299	
Jumlah Tanggungan	C	317	224	56	37	1.1575	0.0006
	B	103	76	16	11	1.0855	

Pada perhitungan tabel node awal dapat diketahui bahwa atribut dengan Gain tertinggi adalah pekerjaan orang tua dengan nilai 0.0120, sehingga atribut pekerjaan orang tua dapat menjadi node akar.

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, dilihat pada Tabel 5.23.

Tabel 5. 23 Hasil *confusion matrix* uji coba 80:20 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	16	0
Tepat	0	57	0
Tidak Bayar	0	11	0
Total	0	84	0

Berdasarkan Tabel 5.23 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+57+0}{84} \times 100\% = 67\%$$

$$Recall = \frac{57}{0+57} \times 100\% = 1\%$$

$$Presisi = \frac{57}{27+57} \times 100\% = 68\%$$

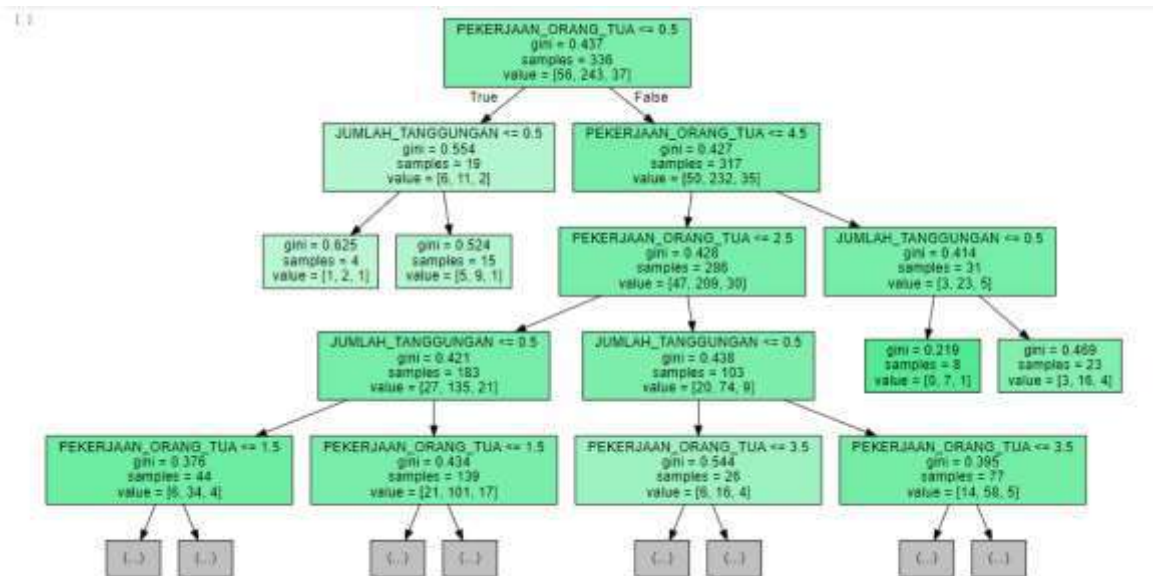
$$F1 - score = \frac{2 \times (1 \times 68)}{1+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode decision tree diperoleh = 67%, presisi= 68%, dan recall = 1%, serta F1-score = 81%. Dapat dilihat pada Tabel 5.24.

Tabel 5. 24 Hasil Uji coba 80:20 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Pekerjaan orang tua 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	67%
			Tepat	0.68	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.4.



Gambar 5. 4 Node akar percobaan 80:20 skenario ketiga

d. Skenario Keempat

Skenario keempat dilakukan dengan 420 dataset, 2 atribut (penghasilan dan jumlah tanggungan). dan terbagi 3 class yaitu terlambat sebanyak 300 data, tepat sebanyak 72 data, dan tidak bayar sebanyak 48 data. Menghitung jumlah dataset keseluruhan dari masing-masing actual class prediksi, terlihat pada tabel 5.25:

Tabel 5. 25 Jumlah dataset uji coba 80:20 skenario keempat

Jumlah Data	Jumlah Terlambat	Jumlah Tepat	Jumlah Tidak Bayar
420	300	72	48

Dari dataset sebanyak 420 dibagi menjadi 2, dengan perbandingan 80:20, data *training* 80% (336 record) sisanya akan dijadikan sebagai data *testing* 20% (84 record). Data kondisi siwa tersebut lalu diklasifikasikan menjadi 2 bagian, yaitu pembagian berdasarkan pekerjaan orang tua dan jumlah tanggungan.

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 55, C (Tepat) = 17, C (Tidak bayar) = 10, hasil jumlah kejadian CT (Terlambat) = 127, CT (Tepat) = 23, CT (Tidak bayar) = 21, hasil jumlah kejadian T (Terlambat) = 117, T (Tepat) = 31, T (Tidak bayar) = 16, sedangkan hasil jumlah kejadian ST (Terlambat) = 1, ST (Tepat) = 1, ST (Tidak bayar) = 1. Hasil klasifikasi dapat di lihat pada Tabel 5.26.

Tabel 5. 26 Klasifikasi penghasilan orang tua uji coba 80:20 skenario keempat

Penghasilan Orang Tua Per Bulan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	55	17	10
CT	127	23	21
T	117	31	16
ST	1	1	1

Klasifikasi parameter pekerjaan orang tua, dari data yang dikumpulkan, diperoleh hasil jumlah kejadian C (Terlambat) = 224, C (Tepat) = 56, C (Tidak bayar) = 37, sedangkan hasil jumlah kejadian B (Terlambat) = 76, B (Tepat) = 16, B (Tidak bayar) = 11. Hasil klasifikasi dapat di lihat pada Tabel 5.27.

Tabel 5. 27 Klasifikasi jumlah tanggungan uji coba 80:20 skenario keempat

Jumlah Tanggungan	Jumlah Kejadian Di Pilih		
	Terlambat	Tepat	Tidak bayar
C	224	56	37
B	76	16	11

Setelah masing-masing klasifikasi kejadian parameter telah diketahui, langkah selanjutnya adalah proses perhitungan entropy dan gain. Hasil entropy dan gain dapat di lihat pada Tabel 5.28.

Tabel 5. 28 Node awal uji coba 80:20 skenario keempat

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain
Total		420	300	72	48	1.1405	
Penghasilan Orang Tua Per Bulan	C	82	55	17	10	1.2272	-0.4887
	CT	171	127	23	21	1.0795	

Atribut		Jumlah	Terlambat	Tepat	Tidak bayar	Entropy	Gain	
	T	164	117	31	16	1.1294		
	ST	3	1	1	1	1.5849		
Jumlah Tanggungan		C	317	224	56	37	1.1575	0.0006
		B	103	76	16	11	1.0855	

Pada perhitungan tabel node awal dapat diketahui bahwa atribut dengan Gain tertinggi adalah jumlah tanggungan dengan nilai 0.0006, sehingga atribut pekerjaan orang tua dapat menjadi node akar.

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.29.

Tabel 5. 29 Hasil *confusion matrix* uji coba 80:20 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	16	0
Tepat	1	56	0
Tidak Bayar	1	10	0
Total	2	82	0

Berdasarkan Tabel 5.29 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+56+0}{84} \times 100\% = 67\%$$

$$Recall = \frac{56}{1+56} \times 100\% = 98\%$$

$$Presisi = \frac{56}{26+56} \times 100\% = 68\%$$

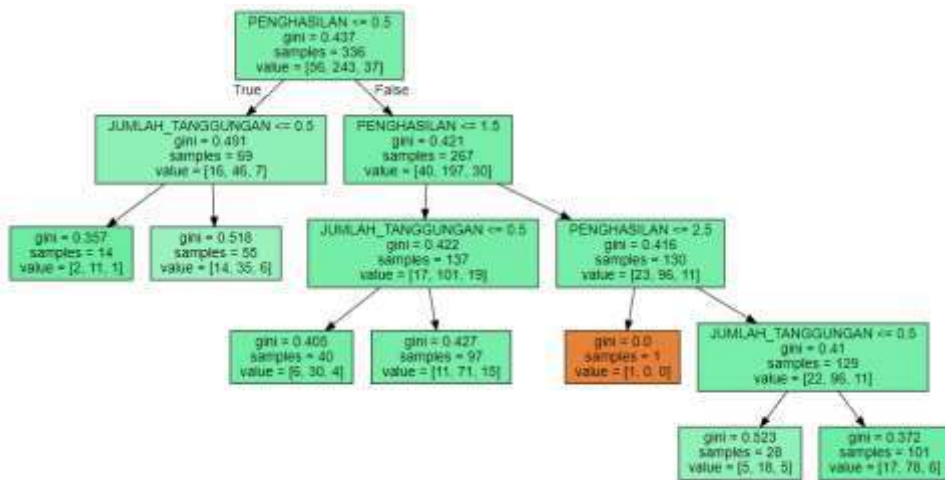
$$F1 - score = \frac{2 \times (98 \times 68)}{98+68} = 81\%$$

Pada hasil perhitungan persamaan di atas, dihasilkan akurasi dengan metode decision tree diperoleh = 67%, presisi= 68%, dan recall = 98%, serta F1-score = 81%, dapat dilihat pada Tabel 5.30.

Tabel 5. 30 Hasil uji coba 80:20 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
80%	20%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.68	0.98	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.5.



Gambar 5. 5 Node akar percobaan 80:20 skenario keempat

5.2.2. Percobaan Kedua Data Uji 75:25

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan kedua dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.31.

Tabel 5. 31 Hasil *confusion matrix* uji coba 75:25 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	2	70	0
Tidak Bayar	1	12	0
Total	3	102	0

Berdasarkan Tabel 4.31 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+70+0}{105} \times 100\% = 67\%$$

$$Presisi = \frac{70}{32+70} \times 100\% = 69\%$$

$$Recall = \frac{70}{2+70} \times 100\% = 97\%$$

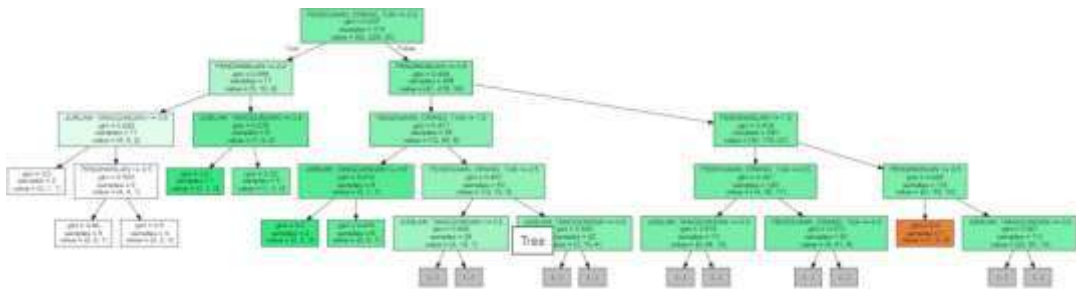
$$F1 - score = \frac{2 \times (97 \times 69)}{97+69} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode decision tree , presisi = 69%, recall= 97% , dan F1-score= 80%., dapat dilihat pada Tabel 5.32.

Tabel 5. 32 Hasil uji coba 75:25 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1. Pekerjaan orang tua 2. Penghasilan 3. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.69	0.97	0.80	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.6.



Gambar 5. 6 Node akar percobaan 75:25 skenario pertama

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan

orang tua dan penghasilan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 record) sisanya akan dijadikan sebagai data *testing* 25% (105 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.33.

Tabel 5. 33 Hasil *confusion matrix* uji coba 75:25 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	1	19	0
Tepat	2	70	0
Tidak Bayar	1	12	0
Total	4	101	0

Berdasarkan Tabel 5.33 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+70+0}{105} \times 100\% = 67\%$$

$$Presisi = \frac{70}{31+70} \times 100\% = 69\%$$

$$Recall = \frac{70}{2+70} \times 100\% = 97\%$$

$$F1 - score = \frac{2 \times (97 \times 69)}{97+69} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 67 % dengan metode decision tree , presisi = 69%, recall= 97% , dan F1-score= 81%, dapat dilihat pada Tabel 5.34.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.35.

Tabel 5. 35 Hasil *confusion matrix* uji coba 75:25 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	0	72	0
Tidak Bayar	0	13	0
Total	0	105	0

Berdasarkan Tabel 5.35 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+72+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{72}{33+0} \times 100\% = 69\%$$

$$Recall = \frac{72}{0+72} \times 100\% = 1\%$$

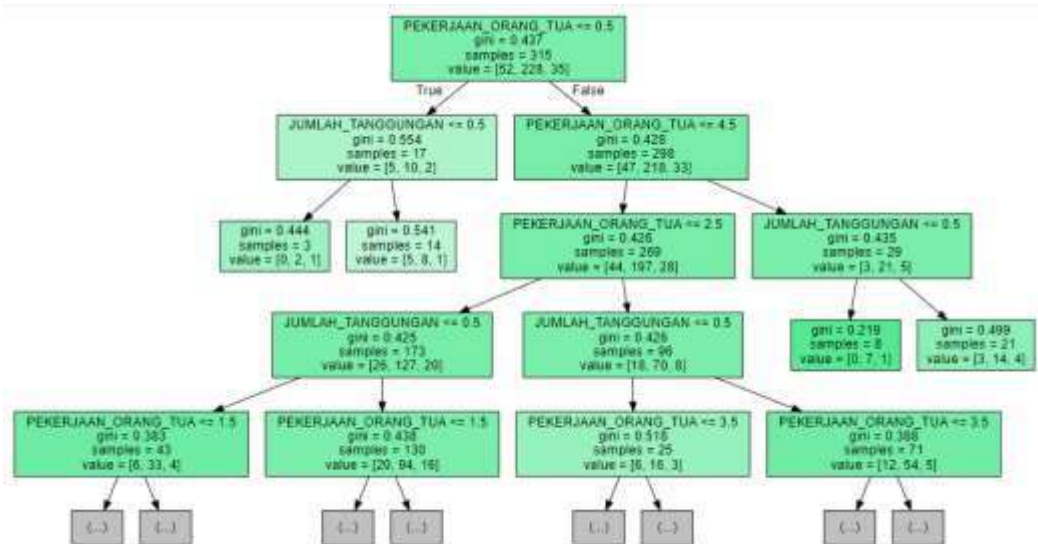
$$F1 - score = \frac{2 \times (1 \times 81)}{1+81} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode decision tree , presisi = 69%, recall= 1% , dan F1-score= 81%, dapat dilihat pada Tabel 5.36.

Tabel 5. 36 Hasil uji coba 75:25 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1. Pekerjaan orang tua 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	68%
			Tepat	0.69	1.00	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.8.



Gambar 5. 8 Node akar percobaan 75:25 skenario ketiga

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 75:25, data *training* 75% (315 *record*) sisanya akan dijadikan sebagai data *testing* 25% (105 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.37.

Tabel 5. 37 Hasil *confusion matrix* uji coba 75:25 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	20	0
Tepat	1	71	0
Tidak Bayar	1	12	0
Total	2	103	0

Berdasarkan Tabel 5.37 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{0+71+0}{105} \times 100\% = 68\%$$

$$Presisi = \frac{71}{32+71} \times 100\% = 69\%$$

$$Recall = \frac{71}{1+71} \times 100\% = 99\%$$

$$F1 - score = \frac{2 \times (99 \times 69)}{99+69} = 81\%$$

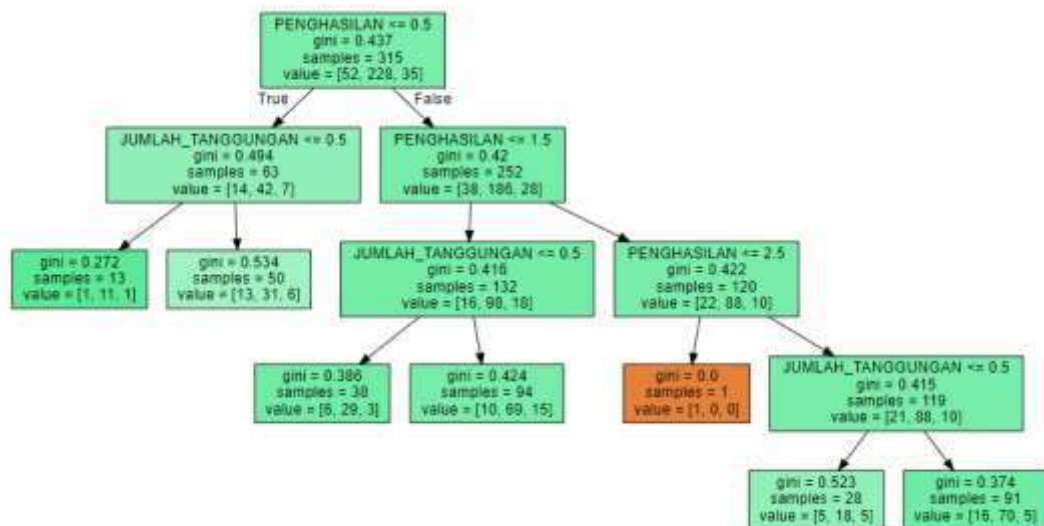
Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode decision tree , presisi = 69%, recall= 99% , dan F1-score= 81%, dapat dilihat pada Tabel 5.38

Tabel 5. 38 Hasil uji coba 75:25 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
75%	25%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	67%
			Tepat	0.69	0.99	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah

actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.9.



Gambar 5. 9 Node akar percobaan 75:25 skenario keempat

5.2.3. Percobaan Ketiga Data Uji 70:30

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan kedua dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 *record*) sisanya akan dijadikan sebagai data *testing* 30% (126 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.39.

Tabel 5. 39 Hasil *confusion matrix* uji coba 70:30 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	3	22	0
Tepat	7	76	1
Tidak Bayar	1	16	0
Total	11	114	1

Berdasarkan Tabel 5.39 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{79}{126} \times 100\% = 62\%$$

$$Presisi = \frac{79}{38+79} \times 100\% = 67\%$$

$$Recall = \frac{79}{8+79} \times 100\% = 90\%$$

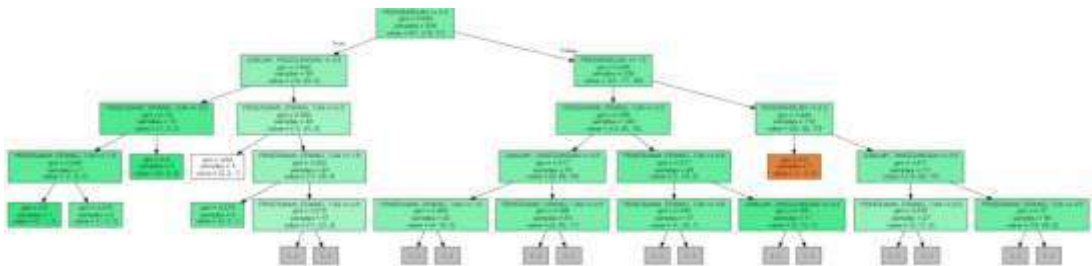
$$F1 - score = \frac{2 \times (90 \times 67)}{90+67} = 77\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 62 % dengan metode decision tree , presisi = 67%, recall= 90% , dan F1-score= 77%, dapat di lihat pada Tabel 5.40.

Tabel 5. 40 Hasil uji coba 70:30 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua 2. Penghasilan 3. Jumlah tanggungan	Terlambat	0.27	0.12	0.17	62%
			Tepat	0.67	0.90	0.77	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.10.



Gambar 5. 10 Node akar percobaan 70:30 skenario pertama

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan), dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 record) sisanya akan dijadikan sebagai data *testing* 30% (126 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.41.

Tabel 5. 41 Hasil *confusion matrix* uji coba 70:30 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	1	24	0
Tepat	2	82	0
Tidak Bayar	1	16	0

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Total	4	122	0

Berdasarkan Tabel 5.41 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{83}{126} \times 100\% = 65\%$$

$$Presisi = \frac{83}{40+70} \times 100\% = 67\%$$

$$Recall = \frac{83}{2+83} \times 100\% = 98\%$$

$$F1 - score = \frac{2 \times (98 \times 67)}{98+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 65 % dengan metode decision tree , presisi = 67%, recall= 98% , dan F1-score= 80%, dilihat pada Tabel 5.42.

Tabel 5. 42 Hasil uji coba 70:30 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua 2. Penghasilan	Terlambat	0.25	0.04	0.07	65%
			Tepat	0.67	0.98	0.80	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.11.

Berdasarkan Tabel 5.43 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{84}{126} \times 100\% = 66\%$$

$$Presisi = \frac{84}{42+84} \times 100\% = 67\%$$

$$Recall = \frac{84}{0+84} \times 100\% = 100\%$$

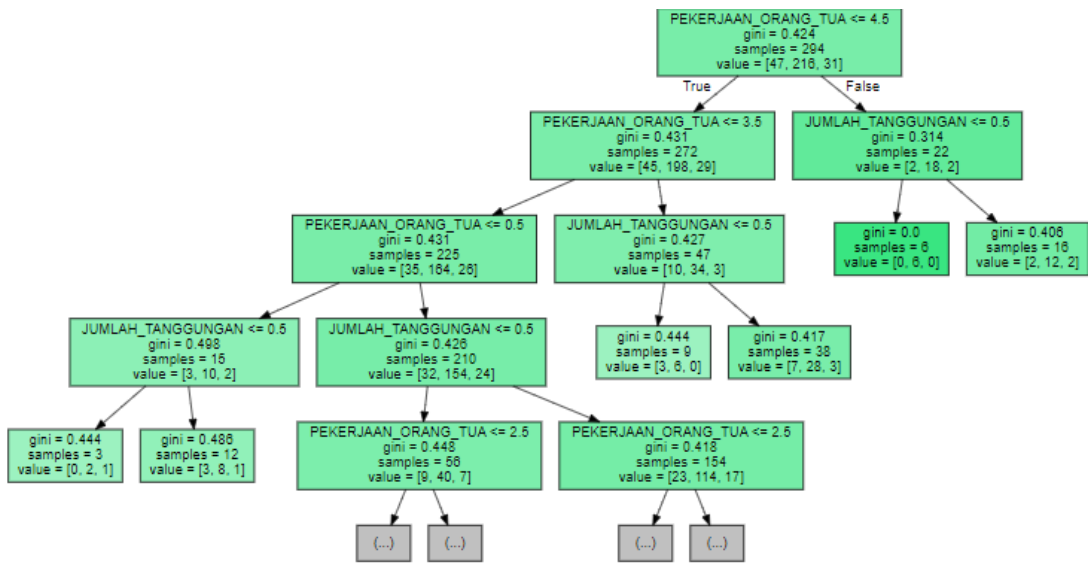
$$F1 - score = \frac{2 \times (1 \times 67)}{1+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 66 % dengan metode decision tree , presisi = 67%, recall= 100% , dan F1-score= 80%, dilihat pada pada Tabel 5.44.

Tabel 5. 44 Hasil uji coba 70:30 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1. Pekerjaan orang tua 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	66%
			Tepat	0.67	1.00	0.80	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.12.



Gambar 5. 12 Node akar percobaan 70:30 skenario ketiga

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 70:30, data *training* 70% (294 record) sisanya akan dijadikan sebagai data *testing* 30% (126 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.45.

Tabel 5. 45 Hasil *confusion matrix* uji coba 70:30 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	25	0
Tepat	1	83	0

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Tidak Bayar	1	16	0
Total	2	124	0

Berdasarkan Tabel 5.45 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{83}{126} \times 100\% = 65\%$$

$$Presisi = \frac{83}{41+83} \times 100\% = 67\%$$

$$Recall = \frac{83}{1+83} \times 100\% = 99\%$$

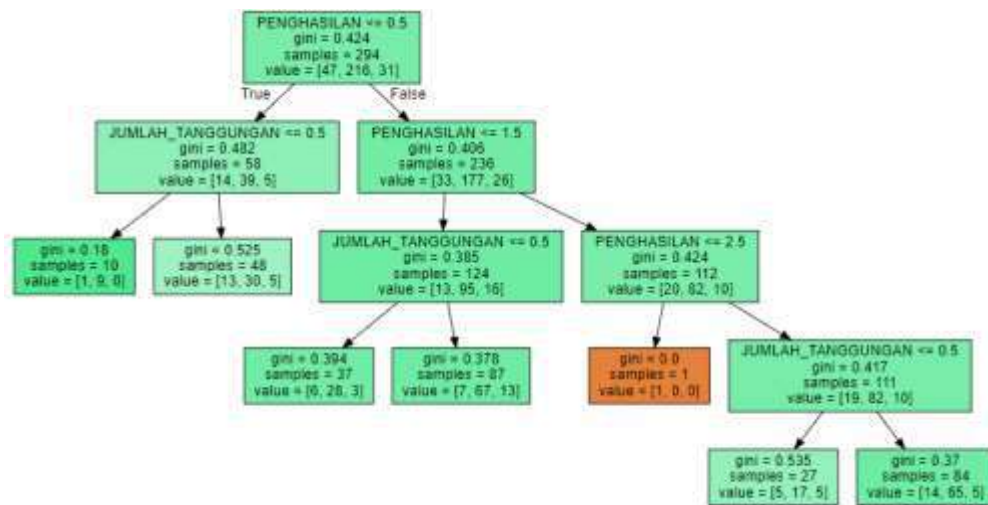
$$F1 - score = \frac{2 \times (99 \times 67)}{99+67} = 80\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 65 % dengan metode decision tree , presisi = 67%, recall= 99% , dan F1-score= 80%, dilihat pada Tabel 5.46.

Tabel 5. 46 Hasil uji coba 70:30 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
70%	30%	1.Penghasilan 2.Jumlah tanggungan	Terlambat	0.00	0.00	0.00	65%
			Tepat	0.67	0.99	0.80	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.13.



Gambar 5. 13 Node akar percobaan 70:30 skenario keempat

5.2.4. Percobaan Keempat Data Uji 60:40

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan keempat dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.47.

Tabel 5. 47 Hasil *confusion matrix* uji coba 60:40 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	3	27	0
Tepat	10	106	1
Tidak Bayar	2	19	0
Total	15	152	1

Berdasarkan Tabel 5.47 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{109}{168} \times 100\% = 64\%$$

$$Presisi = \frac{109}{46+109} \times 100\% = 70\%$$

$$Recall = \frac{109}{11+109} \times 100\% = 91\%$$

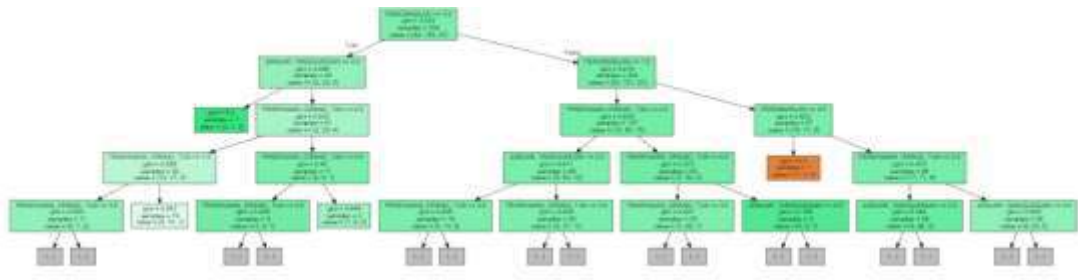
$$F1 - score = \frac{2 \times (91 \times 70)}{91+70} = 79\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 62 % dengan metode decision tree , presisi = 67%, recall= 90% , dan F1-score= 77%, dilihat pada Tabel 5.48.

Tabel 5. 48 Hasil uji coba 60:40 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Pekerjaan orang tua	Terlambat	0.20	0.10	0.13	64%
		2. Penghasilan	Tepat	0.70	0.91	0.79	
		3. Jumlah tanggungan	Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.14.



Gambar 5. 14 Node akar percobaan 60:40 skenario pertama

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan), dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 record) sisanya akan dijadikan sebagai data *testing* 40% (168 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.49.

Tabel 5. 49 Hasil *confusion matrix* uji coba 60:40 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	1	29	0
Tepat	2	115	0
Tidak Bayar	1	20	0
Total	4	164	0

Berdasarkan Tabel 5.49 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 *record*) sisanya akan dijadikan sebagai data *testing* 40% (168 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.51.

Tabel 5. 51 Hasil *confusion matrix* uji coba 60:40 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	0	117	0
Tidak Bayar	0	21	0
Total	0	168	0

Berdasarkan Tabel 5.51 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{117}{168} \times 100\% = 69\%$$

$$Presisi = \frac{117}{51+117} \times 100\% = 70\%$$

$$Recall = \frac{117}{0+117} \times 100\% = 100\%$$

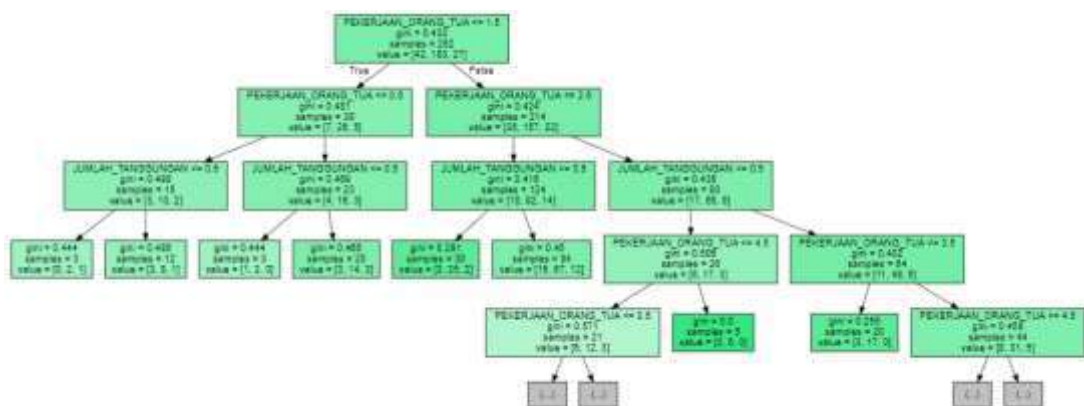
$$F1 - score = \frac{2 \times (1 \times 70)}{1+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode decision tree , presisi = 70%, recall= 1% , dan F1-score= 82%, dilihat pada Tabel 5.52.

Tabel 5. 52 Hasil uji coba 60:40 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Pekerjaan orang tua 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	69%
			Tepat	0.70	1.00	0.82	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.16.



Gambar 5. 16 Node akar percobaan 60:40 skenario ketiga

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 60:40, data *training* 60% (252 record) sisanya akan dijadikan sebagai data *testing* 40% (168 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.53.

Tabel 5. 53 Hasil *confusion matrix* uji coba 60:40 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	30	0
Tepat	1	116	0
Tidak Bayar	1	20	0
Total	2	166	0

Berdasarkan Tabel 5.53 tentang hasil *confusion matrix* metode naive bayes maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{116}{168} \times 100\% = 69\%$$

$$Presisi = \frac{116}{50+116} \times 100\% = 70\%$$

$$Recall = \frac{116}{1+116} \times 100\% = 99\%$$

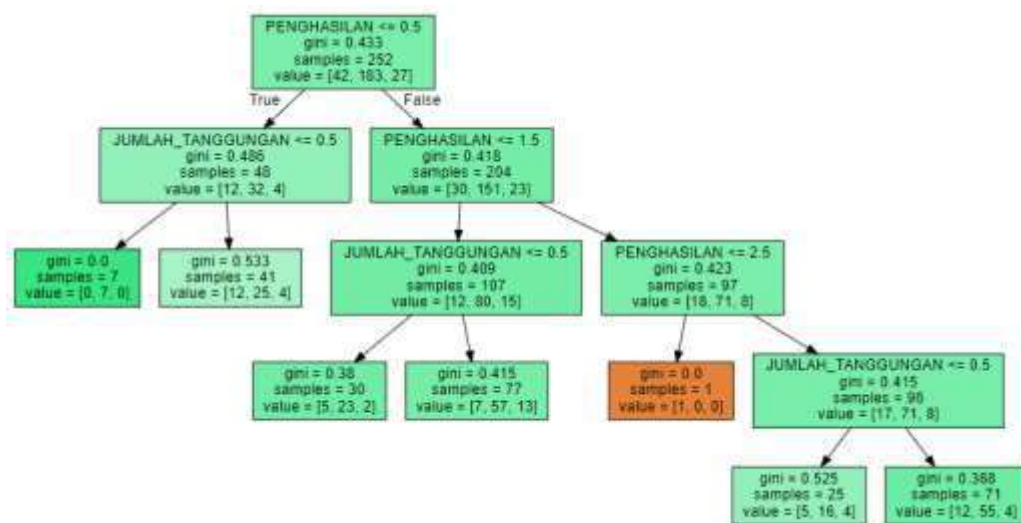
$$F1 - score = \frac{2 \times (99 \times 70)}{99+70} = 82\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 69 % dengan metode decision tree , presisi = 70%, recall= 99% , dan F1-score= 82%, dilihat pada Tabel 5.54.

Tabel 5. 54 Hasil uji coba 60:40 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
60%	40%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	69%
			Tepat	0.70	0.99	0.82	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.17.



Gambar 5. 17 Node akar percobaan 60:40 skenario keempat

5.2.5. Percobaan Kelima Data Uji 50:50

Pada percobaan ini, data dan atribut yang digunakan adalah data yang sama dengan percobaan pertama. Langkah-langkah yang dilakukan sama dengan percobaan pertama dengan menghitung jumlah dataset dan probabilitas. Berikut percobaan keempat dilakukan menjadi 4 skenario, yaitu :

a. Skenario Pertama

Skenario pertama dilakukan dengan langkah yang sama pada skenario pertama di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 3 atribut (pekerjaan orang tua, penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 50:50, data *training* 50% (210 *record*) sisanya akan dijadikan sebagai data *testing* 50% (210 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.55.

Tabel 5. 55 Hasil *confusion matrix* uji coba 50:50 skenario pertama

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	2	30	2
Tepat	32	118	1
Tidak Bayar	3	22	0
Total	37	170	3

Berdasarkan Tabel 5.55 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{120}{210} \times 100\% = 57\%$$

$$Presisi = \frac{120}{52+120} \times 100\% = 69\%$$

$$Recall = \frac{120}{32+120} \times 100\% = 78\%$$

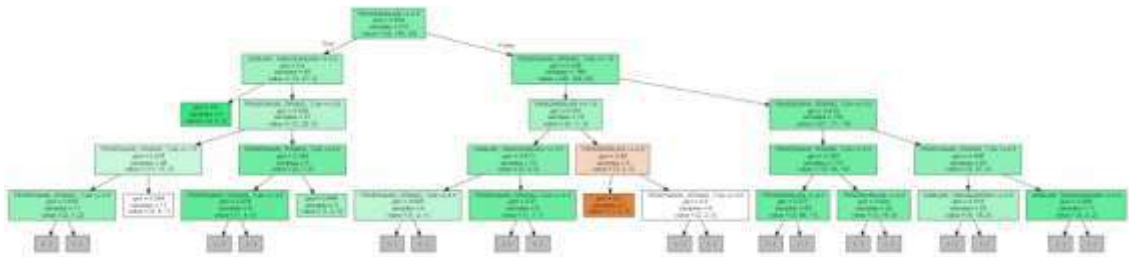
$$F1 - score = \frac{2 \times (69 \times 78)}{69+78} = 74\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 57 % dengan metode decision tree , presisi = 69%, recall= 78% , dan F1-score= 74%, dilihat pada Tabel 5.56.

Tabel 5. 56 Hasil uji coba 50:50 skenario pertama

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua 2. Penghasilan 3. Jumlah tanggungan	Terlambat	0.05	0.06	0.06	57%
			Tepat	0.69	0.78	0.74	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.18.



Gambar 5. 18 Node akar percobaan 50:50 skenario pertama

b. Skenario Kedua

Skenario kedua dilakukan dengan langkah yang sama pada skenario kedua di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan penghasilan), dataset dibagi menjadi 2, dengan perbandingan 50:50, data *training* 50% (210 record) sisanya akan dijadikan sebagai data *testing* 50% (210 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai

performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.57.

Tabel 5. 57 Hasil *confusion matrix* uji coba 50:50 skenario kedua

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	2	32	0
Tepat	18	133	0
Tidak Bayar	1	24	0
Total	21	189	0

Berdasarkan Tabel 5.57 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{135}{210} \times 100\% = 64\%$$

$$Presisi = \frac{135}{56+135} \times 100\% = 70\%$$

$$Recall = \frac{135}{18+135} \times 100\% = 88\%$$

$$F1 - score = \frac{2 \times (88 \times 70)}{88+70} = 78\%$$

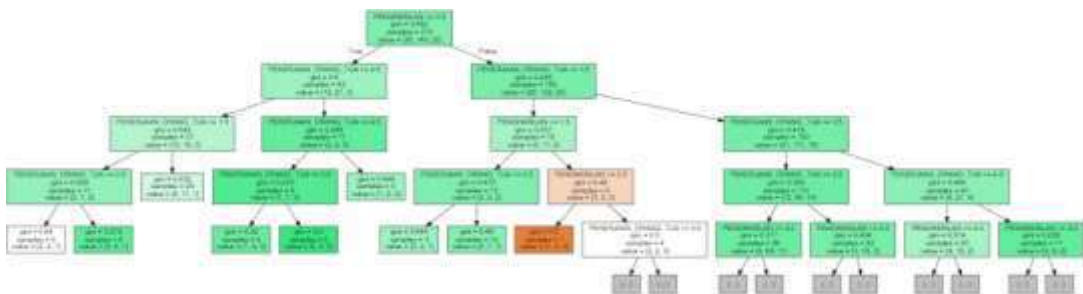
Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 64 % dengan metode decision tree , presisi = 70%, recall= 88% , dan F1-score= 78%, dilihat pada Tabel 5.58.

Tabel 5. 58 Hasil uji coba 50:50 skenario kedua

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua	Terlambat	0.10	0.06	0.07	64%
			Tepat	0.70	0.88	0.78	

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
		2. Penghasilan	Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dilihat pada gambar 5.19.



Gambar 5. 19 Node akar percobaan 50:50 skenario kedua

c. Skenario Ketiga

Skenario ketiga dilakukan dengan langkah yang sama pada skenario ketiga di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (pekerjaan orang tua dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 50:50, data *training* 50% (210 record) sisanya akan dijadikan sebagai data *testing* 50% (210 record).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.59.

Tabel 5. 59 Hasil *confusion matrix* uji coba 50:50 skenario ketiga

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	34	0
Tepat	8	143	0
Tidak Bayar	1	24	0
Total	9	201	0

Berdasarkan Tabel 5.59 tentang hasil *confusion matrix* metode decision treemaka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{143}{210} \times 100\% = 68\%$$

$$Presisi = \frac{143}{58+143} \times 100\% = 71\%$$

$$Recall = \frac{143}{8+143} \times 100\% = 95\%$$

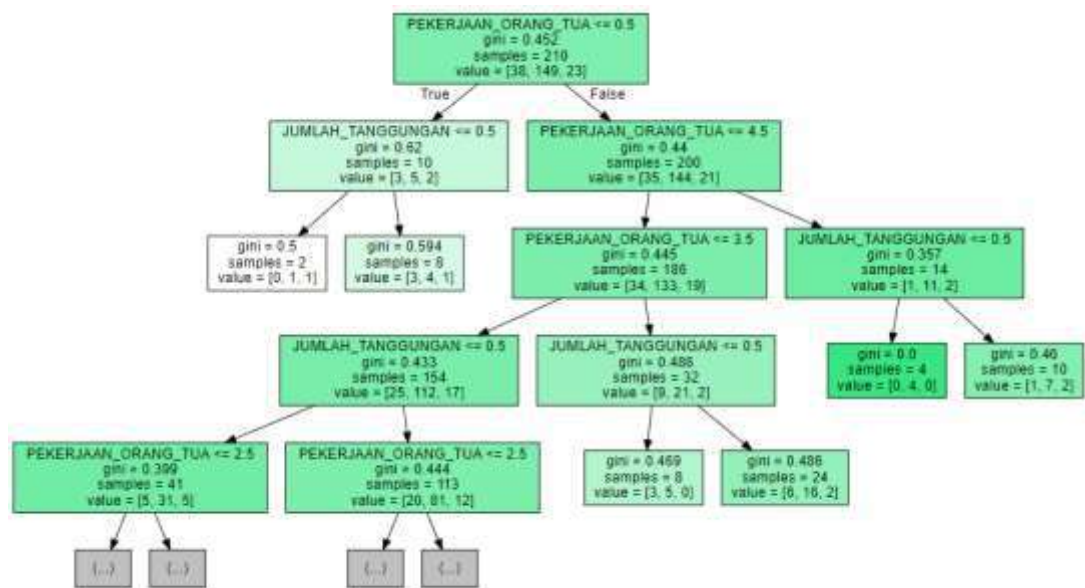
$$F1 - score = \frac{2 \times (71 \times 95)}{71+95} = 81\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 68 % dengan metode decision tree , presisi = 71%, recall= 95% , dan F1-score= 81%, dapat dilihat pada Tabel 5.60.

Tabel 5. 60 Hasil uji coba 50:50 skenario ketiga

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Pekerjaan orang tua	Terlambat	0.00	0.00	0.00	68%
		2. Jumlah tanggungan	Tepat	0.71	0.95	0.81	
			Tidak Bayar	0.00	0.00	0.00	

Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.20.



Gambar 5. 20 Node akar percobaan 50:50 skenario ketiga

d. Skenario Keempat

Skenario keempat dilakukan dengan langkah yang sama pada skenario keempat di percobaan pertama, yaitu dari dataset sebanyak 420, dengan 2 atribut (penghasilan dan jumlah tanggungan), dataset dibagi menjadi 2, dengan perbandingan 50:50, data *training* 50% (210 *record*) sisanya akan dijadikan sebagai data *testing* 50% (210 *record*).

Dari hasil proses klasifikasi proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree dengan cara menentukan nilai akurasi, presisi, recall dan F1-score. Dengan menggunakan *confusion matrix* nilai performa dari metode decision tree dapat diketahui. Dari *confusion matrix* kita dapat mengetahui klasifikasi jumlah data uji yang benar dan yang salah.

Dari hasil proses klasifikasi tentang hasil uji coba dilakukan proses selanjutnya yaitu untuk mengetahui performa dari metode klasifikasi decision tree yang ditampilkan pada hasil *confusion matrix*, ditunjukkan pada Tabel 5.61.

Tabel 5. 61 Hasil *confusion matrix* uji coba 50:50 skenario keempat

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Terlambat	0	34	0

Correct Classification	Classification		
	Terlambat	Tepat	Tidak bayar
Tepat	1	150	0
Tidak Bayar	1	24	0
Total	2	208	0

Berdasarkan Tabel 5.61 tentang hasil *confusion matrix* metode decision tree maka dilakukan pengujian sesuai rumus.

Hasil pengujian :

$$Akurasi = \frac{150}{210} \times 100\% = 71\%$$

$$Presisi = \frac{150}{58+150} \times 100\% = 72\%$$

$$Recall = \frac{150}{1+150} \times 100\% = 99\%$$

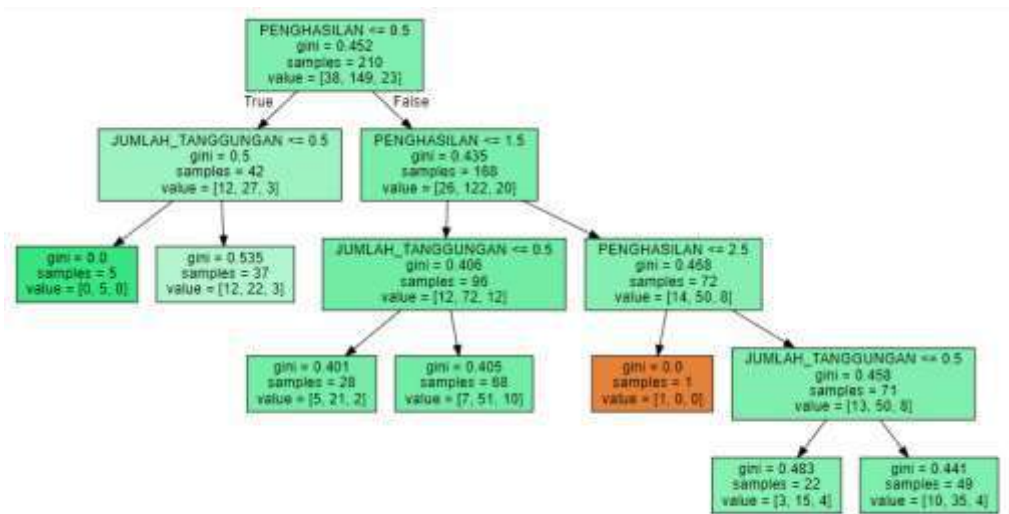
$$F1 - score = \frac{2 \times (99 \times 72)}{99+72} = 84\%$$

Pada hasil perhitungan *confusion matrix*, dihasilkan akurasi = 71 % dengan metode decision tree , presisi = 72%, recall= 99% , dan F1-score= 84%, dapat dilihat pada Tabel 5.62.

Tabel 5. 62 Hasil uji coba 50:50 skenario keempat

Perbandingan Data		Atribut	Correct Classification	Hasil Klasifikasi			Akurasi
Training	Testing			Presisi	Recall	F1-Score	
50%	50%	1. Penghasilan 2. Jumlah tanggungan	Terlambat	0.00	0.00	0.00	71%
			Tepat	0.72	0.99	0.84	
			Tidak Bayar	0.00	0.00	0.00	

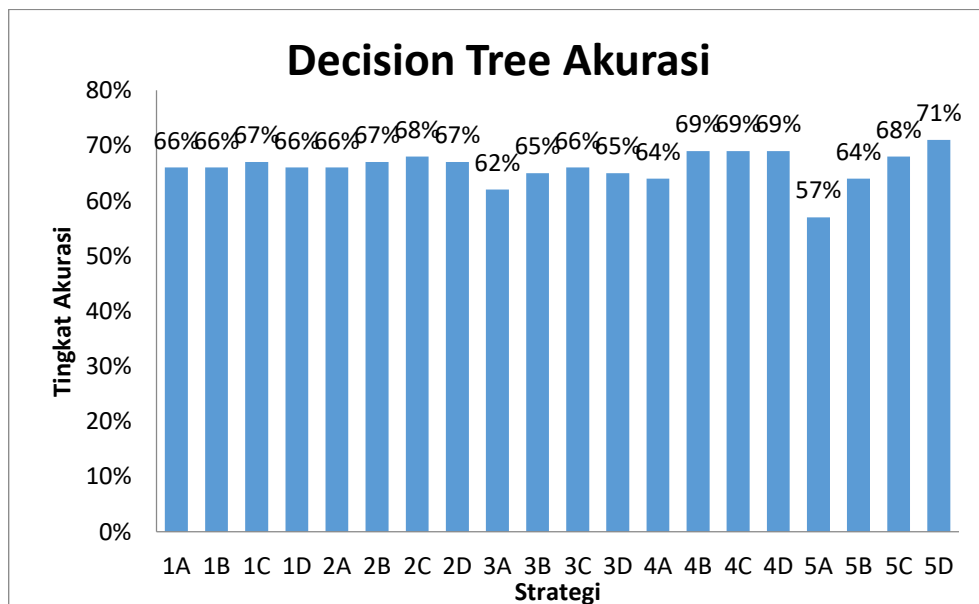
Berdasarkan perhitungan kemungkinan actual class terlambat, tepat dan tidak bayar, maka dapat di hasilkan prediksi pada data baru yang di inputkan adalah actual class terlambat yang memiliki nilai paling tinggi. Hasil node akar dapat dilihat pada gambar 5.21.



Gambar 5. 21 Node akar percobaan 50:50 skenario keempat

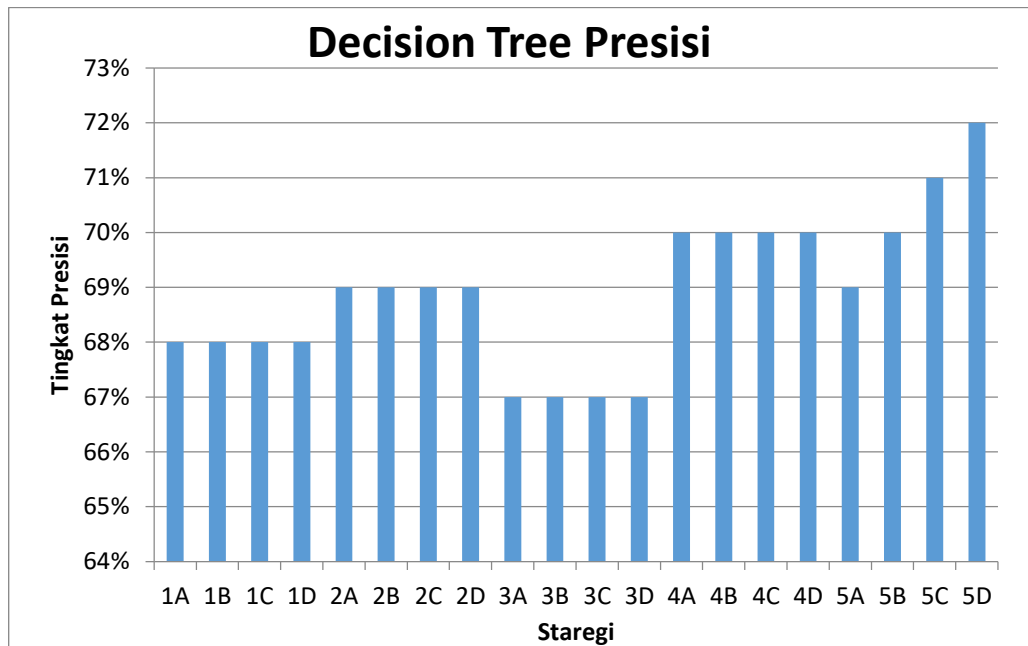
5.3. Kesimpulan

Setelah melakukan uji coba sebanyak 5 percobaan prediksi , dengan masing-masing percobaan terdapat 4 skenario menggunakan metode decision tree, maka dapat kita simpulkan pada Grafik 5.1.



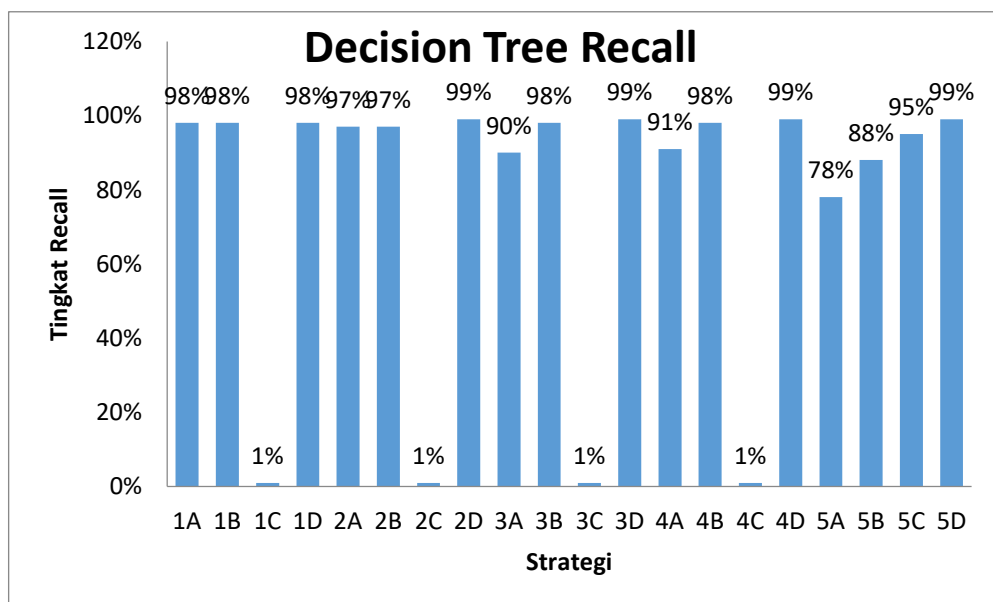
Grafik 5. 1 Hasil uji coba akurasi untuk semua skenario

Dari Grafik 5.1, di jelaskan bahwasannya akurasi tertinggi terdapat pada skenario 5D dengan nilai akurasi :71%.



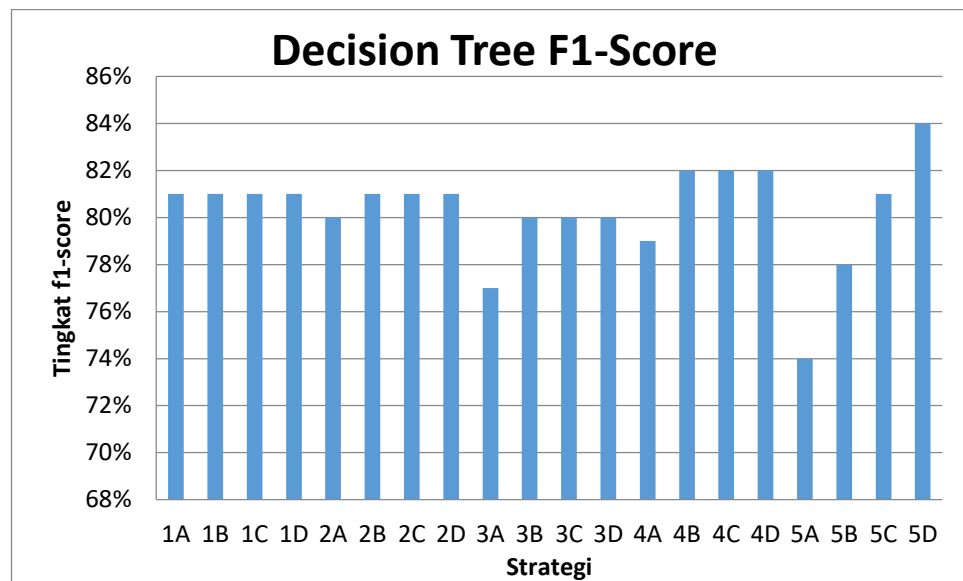
Grafik 5. 2 Hasil uji coba presisi untuk semua skenario

Dari hasil Grafik 5.2, dijelaskan bahwasannya hasil presisi tertinggi terdapat pada skenario 5D dengan nilai presisi: 72%.



Grafik 5. 3 Hasil uji coba recall untuk semua skenario

Dari hasil Grafik 5.3, dijelaskan bahwasannya hasil recall tertinggi terdapat pada skenario 2D, 3D, 4D dan 5D dengan nilai recall : 99%.



Grafik 5. 4 Hasil uji coba f1-score untuk semua skenario

Dari hasil Grafik 5.4, dijelaskan bahwasannya hasil f1-score tertinggi terdapat pada skenario 5D dengan nilai f1-score: 84%..

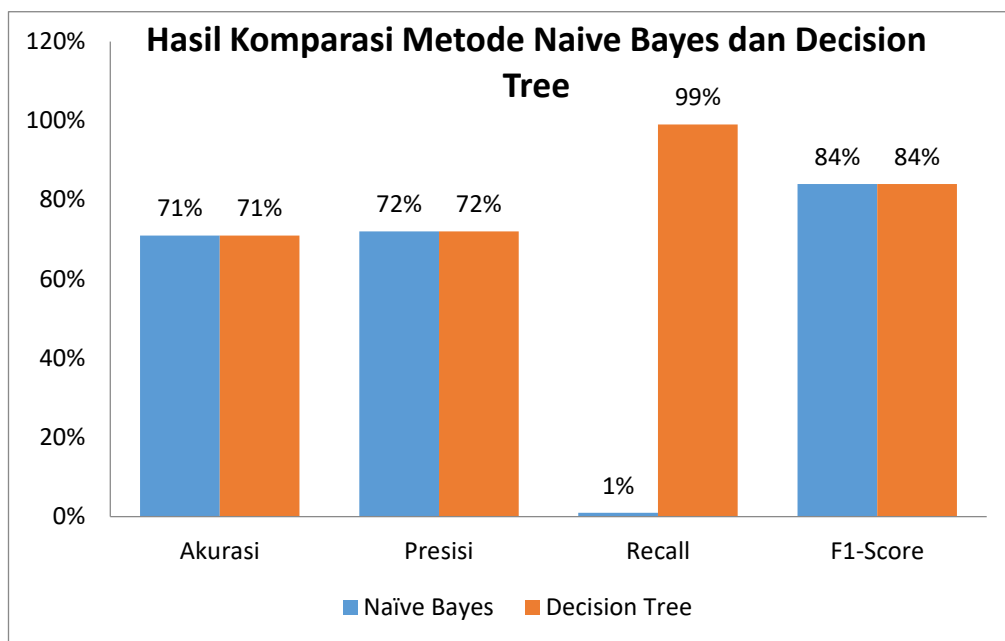
Dari grafik 5.1 sampai 5.4 yang ditampilkan, maka dapat di simpulkan skenario 5D menghasilkan akurasi tertinggi yaitu 71%, presisi = 72%, recall= 99% , dan F1-score= 84%, dengan menggunakan 2 parameter (penghasilan orang tua dan jumlah tanggungan orang tua).

Jadi kesimpulannya pada penggunaan metode decision tree, 2, melakukan perbandingan data uji 50:50, dengan menggunakan 2 parameter (penghasilan orang tua dan jumlah tanggungan orang tua) memiliki pengaruh dalam mengklasifikasi ekonomi siswa.

BAB VI PEMBAHASAN

6.1. Pembahasan Komparasi Algoritma

Pada bab ini dilakukan komparasi atau perbandingan hasil performa dari algoritma naïve bayes dengan algoritma decision tree, dimana alat ukur yang di gunakan untuk membandingkan adalah nilai akurasi, presisi, recall dan F1-score, yang disajikan dalam bentuk Grafik 6.1.



Grafik 6. 1 Hasil komparasi algoritma naïve bayes dan decision tree

Berdasarkan Grafik 6.1, maka dapat ketahui tingkat accuracy pada perhitungan menggunakan python dengan data testing sebanyak 210 menggunakan metode naïve bayes dan menggunakan semua parameter yaitu pekerjaan orang tua, penghasilan orang tua dan jumlah tanggungan orang tua, menghasilkan akurasi sebesar 71%, kelas tepat presisi = 72%, recall= 1% , dan F1-score= 84%, kelas terlambat presisi = 0%, recall= 0% , dan F1-score= 0%, dan kelas tidak bayar presisi = 0%, recall= 0% , dan F1-score= 0%.

Sedangkan dengan menggunakan metode decision tree dan menggunakan dua parameter yaitu penghasilan orang tua dan jumlah tanggungan menghasilkan akurasi = 71%, kelas tepat presisi = 72%, recall= 99% , dan F1-score= 84%, kelas

terlambat presisi = 0%, recall= 0% , dan F1-score= 0%, dan kelas tidak bayar presisi = 0%, recall= 0% , dan F1-score= 0%.

Faktor- faktor yang berpengaruh dalam mengklasifikasi ekonomi siswa sekolah dilihat pada Tabel 6.1.

Tabel 6. 1 Hasil rasio pekerjaan orang tua

Pekerjaan	Benar	Salah	Total	Persentase Benar	Persentase Salah	Rasio
G	3	6	9	33.33%	66.67%	0.50
PN	11	7	18	61.11%	38.89%	1.57
P	13	8	21	61.90%	38.10%	1.63
W	20	8	28	71.43%	28.57%	2.50
S	77	28	105	73.33%	26.67%	2.75
T	20	7	27	74.07%	25.93%	2.86

Pada Tabel 6.1 dijelaskan dari parameter pekerjaan orang tua didapatkan hasil rasio kesalahan tertinggi, yang artinya banyak terlambat yaitu pekerjaan sebagai guru sebesar 50%, pekerjaan sebagai pegawai negeri, rasio terlambat sebesar 1.57%, pekerjaan sebagai petani, rasio terlambat sebesar sebesar 1.63%, pekerjaan sebagai wiraswasta, rasio terlambat sebesar sebesar 2,5%, pekerjaan swasta, rasio terlambat sebesar sebesar 2,75%, dan pekerjaan sebagai TNI yang memiliki tingkat rasio terlambat terkecil yaitu sebesar 2.86%.

Tabel 6. 2 Hasil rasio penghasilan orang tua

Penghasilan	Benar	Salah	Total	Persentase Benar	Persentase Salah	Rasio
C	57	27	84	67.86%	32.14%	2.11
T	57	27	84	67.86%	32.14%	2.11
CT	63	22	85	74.12%	25.88%	2.86

Pada Tabel 6.2 dijelaskan dari parameter penghasilan orang tua didapatkan hasil rasio kesalahan tertinggi, yang artinya banyak terlambat yaitu orang tua dengan penghasilan cukup sebesar 2.11%, penghasilan dengan gaji tinggi, rasio terlambat sebesar 2.11%, penghasilan cukup tinggi, memiliki rasio terlambat terkecil yaitu sebesar 2.86%.

Tabel 6. 3 Hasil rasio jumlah tanggungan orang tua

Jumlah Tanggungan	Benar	Salah	Total	Persentase Benar	Persentase Salah	Rasio
C	111	49	160	69.38%	30.63%	2.27
B	34	16	50	68.00%	32.00%	2.13

Pada Tabel 6.3 dijelaskan dari parameter jumlah tanggungan orang tua didapatkan hasil rasio kesalahan tertinggi, yang artinya banyak terlambat yaitu orang tua dengan jumlah tanggungan cukup sebesar 2.27%, sedangkan orang tua yang jumlah tanggungannya banyak, memiliki rasio keterlambatan paling kecil yaitu sebesar 2.13%.

Jika rasio keterlambatan dilakukan dengan menggunakan 2 parameter yaitu penghasilan orang tua dan jumlah tanggungan orang tua, maka dihasilkan tingkat rasio keterlambatan pada Tabel 6.4.

Kriteria Penghasilan Dan Jumlah Tanggungan	Benar	Salah	Total	Persentase Benar	Persentase Salah	Rasio
C C	19	14	33	57.58%	42.42%	1.36
T B	10	6	16	62.50%	37.50%	1.67
T C	47	21	68	69.12%	30.88%	2.24
CT B	18	8	26	69.23%	30.77%	2.25
C B	6	2	8	75.00%	25.00%	3.00
CT C	45	14	59	76.27%	23.73%	3.21

Pada Tabel 6.4 dijelaskan dari parameter penghasilan orang tua dan jumlah tanggungan orang tua didapatkan hasil rasio kesalahan tertinggi, yang artinya banyak terlambat yaitu orang tua dengan penghasilannya cukup dan jumlah tanggungan orang tua cukup sebesar 1.36%, penghasilannya tinggi dan jumlah tanggungan orang tua banyak, sebesar 1.67%, orang tua penghasilannya tinggi dan jumlah tanggungan orang tua cukup, sebesar 2.24%, orang tua penghasilannya cukup tinggi dan jumlah tanggungan orang tua banyak, sebesar 2.25, orang tua penghasilannya cukup dan jumlah tanggungan orang tua banyak, sebesar 3%,

sedangkan orang tua penghasilannya cukup tinggi dan jumlah tanggungan orang tua cukup, memiliki rasio keterlambatan terkecil sebesar 3.21%.

Dari tabel hasil rasio keterlambatan maka faktor- faktor yang berpengaruh dalam mengklasifikasi ekonomi siswa sekolah pekerjaan orang tua sebagai guru dengan penghasilan cukup dan jumlah tanggungan cukup beresiko terlambat dalam membayar biaya pendidikan sekolah.

Untuk hasil prediksi menggunakan 210 data pada data testing. Terdapat data yang cocok antara kelas tepat dan prediksi tepat yaitu sebanyak 1 data (true positive), untuk data kelas terlambat yang cocok dengan prediksi terlambat sebanyak 144 data (true negative), untuk data kelas tidak bayar yang cocok dengan prediksi tidak bayar sebanyak 0 data (false negative), data untuk kelas tepat tetapi prediksi terlambat sebanyak 38 data (false negative), data untuk kelas tepat tetapi prediksi tidak bayar sebanyak 0 data (false negative), data untuk kelas terlambat tetapi prediksi tepat sebanyak 1 data (false postive), data untuk kelas terlambat tetapi prediksi tidak bayar sebanyak 0 data (false negative), data untuk kelas tidak bayar tetapi prediksi tepat sebanyak 0 data (false postive) dan data untuk kelas tidak bayar tetapi prediksi terlambat sebanyak 26 data (false negative).

Untuk memperkuat hasil keputusan algoritma yang terbaik, peneliti juga melakukan analisa terhadap hasil perhitungan confusion matrik dari performa terbaik pada skenario percobaan masing-masing algoritma seperti terlihat pada Lampiran 2, tentang komparasi kesesuaian hasil prediksi pada algoritma naïve bayes dan decision tree.

Integrasi yang diharapkan antara ilmu agama Islam dengan sains dan teknologi bukan dipahami dengan memberikan materi pendidikan agama Islam yang diselingi dengan dengan materi sains dan teknologi. Akan tetapi yang dimaksudkan adalah adanya integrasi yang sebenarnya, di mana ketika kita menjelaskan tentang suatu materi ilmu agama Islam dapat didukung oleh fakta sains dan teknologi. Sebab, di dunia yang demikian modern ini, peserta didik

tidak mau hanya sekedar menerima secara dogmatis saja setiap materi pelajaran agama yang mereka terima.

Berikut adalah integrasi Islam yang berhubungan dengan pembiayaan pendidikan, Pembiayaan pendidikan pada dasarnya lahir dari pengertian yaitu “saya percaya” atau “saya menaruh kepercayaan”. Perkataan pembiaya pendidikan yang artinya kepercayaan yang berarti walimurid menaruh kepercayaan kepada seseorang untuk melaksanakan amanah yang diberika oleh walimurid selaku shahibul maal. Dana tersebut harus digunakan dengan benar, adil, dan harus disertai dengan ikatan dan syarat-syarat yang jelas, sebagaimana firman Allah SWT dalam: Q.S. An-Nisa“ (4) ayat 29:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ وَلَا تَقْتُلُوا أَنْفُسَكُمْ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا ۚ ٢٩

Artinya: “Hai orang-orang yang beriman, janganlah kamu saling memakan harta sesamamu dengan jalan yang batil, kecuali dengan jalan perniagaan yang berlaku dengan suka sama suka di antara kamu. Dan janganlah kamu membunuh dirimu; sesungguhnya Allah adalah Maha Penyayang kepadamu”. (Q.S. An-Nisa: 29)

Pada ayat tersebut menurut tafsir Al-Jalalain “(Hai orang-orang yang beriman! Janganlah kamu makan harta sesamamu dengan jalan yang batil) artinya jalan yang haram menurut agama seperti riba dan gasab/merampas (kecuali dengan jalan) atau terjadi (secara perniagaan) menurut suatu qiraat dengan baris di atas sedangkan maksudnya ialah hendaklah harta tersebut harta perniagaan yang berlaku (dengan suka sama suka di antara kamu) berdasar kerelaan hati masing-masing, maka bolehlah kamu memakannya. (Dan janganlah kamu membunuh dirimu) artinya dengan melakukan hal-hal yang menyebabkan kecelakaannya bagaimana pun juga cara dan gejalanya baik di dunia dan di akhirat. (Sesungguhnya Allah Maha Penyayang kepadamu) sehingga dilarang-Nya kamu berbuat demikian”. Ayat tersebut menyuruh kita untuk menginfakkan dan membelanjakan harta kita semata-mata untuk mencari keridhoan Allah Swt. Salah satunya pada bidang pendidikan dengan biaya pendidikan harus direncanakan dan mencari keridhoan Allah SWT.

Menurut surat al-Mujadilah ayat 12–13, Allah SWT secara implisit telah menyebutkan masalah pembiayaan pendidikan:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا تَجَاجَيْتُمُ الرَّسُولَ فَقَدِّمُوا بَيْنَ يَدَيْ نَجْوٰكُمْ صَدَقَةً ۚ ذٰلِكَ خَيْرٌ لَّكُمْ وَأَطْهَرٌ ۚ فَإِن لَّمْ تَجِدُوا فَإِنَّ اللَّهَ غَفُورٌ رَّحِيمٌ (١٢)

ءَأَشْفَقْتُمْ أَن تُقَدِّمُوا بَيْنَ يَدَيْ نَجْوٰكُمْ صَدَقَاتٍ ۚ فَإِذ لَّمْ تَفْعَلُوا وَتَابَ اللَّهُ عَلَيْكُمْ فَأَقْبِمُوا الصَّلَاةَ وَآتُوا الزَّكَاةَ وَأَطِيعُوا اللَّهَ وَرَسُولَهُ ۗ وَاللَّهُ خَبِيرٌ بِمَا تَعْمَلُونَ (١٣)

Artinya adalah: "Wahai orang-orang yang beriman, sebelum melakukan percakapan khusus dengan Rasulullah, Anda harus memberi sedekah kepada orang miskin. Itu lebih bersih dan lebih baik untuk Anda. Namun, sesungguhnya Allah Maha Pengampun lagi Maha Penyayang jika kamu tidak menerima (yang akan disedekahkan). Apakah Anda khawatir menjadi miskin karena Anda memberi uang sebelum berbicara dengan Rasul? Namun, jika Anda gagal lakukan, dan Tuhan menyediakan Anda pengampunan, maka berdoa, berzakat, dan ikutilah Allah dan Rasul-Nya! Dan Allah mengetahui segala sesuatu tentang apa yang kamu kerjakan".

Ayat tersebut menurut tafsir Al-Jalalain "Hai orang-orang yang beriman, jika hendak mengadakan pembicaraan khusus dengan rasul, hendaknya mengeluarkan sebelum pembicaraan (sedekah) kalian sebelum diadakan pembicaraan khusus. Itu lebih baik bagi kalian dan lebih bersih, dan jika kalian tidak menemukan apa Anda bersedekah, Allah akan mengampuni Anda untuk pembicaraan khusus yang akan Anda berikan (sekali lagi, Allah Maha Penyayang untuk Anda). Maksudnya adalah bahwa berbicara khusus itu meskipun Anda tidak memberi sedekah bukanlah dosa". Di dalam ayat tersebut kita diajarkan untuk bermusyawarah untuk menentukan biaya pendidikan yang direncanakan, oleh karena itu wali murid, komite dan pihak sekolah harus melakukan musyawarah untuk merencanakan biaya pendidikan supaya semua berjalan dengan lancar dan ikhlas di hati para wali murid.

Ketika Allah menyuruh kita menginfakkan harta kita ke jalan yang benar, hal tersebut tersirat dalam surah Ali Imrān ayat: 14 :

رُيِّنَ لِلنَّاسِ حُبُّ الشَّهَوَاتِ مِنَ النِّسَاءِ وَالْبَنِينَ وَالْقَنَاطِيرِ الْمُقَنْطَرَةِ مِنَ الذَّهَبِ وَالْفِضَّةِ وَالْخَيْلِ الْمُسَوَّمَةِ وَالْأَنْعَامِ وَالْحَرْثِ ۗ ذٰلِكَ مَتَاعُ الدُّنْيَا ۗ وَاللَّهُ عِنْدَهُ حُسْنُ الْمَاٰبِ (١٤)

Artinya : Dijadikan indah pada (pandangan) manusia kecintaan kepada apa-apa yang diingini, yaitu: wanita-wanita, anak-anak, harta yang banyak dari jenis emas, perak, kuda pilihan, binatang-binatang ternak dan sawah ladang. Itulah kesenangan hidup di dunia, dan di sisi Allah-lah tempat kembali yang baik (surga). (QS. Ali Imrān :14)

Ayat tersebut di tafsirkan dari tafsir Al- Jalalain yakni “(Dijadikan indah pada pandangan manusia kecintaan kepada syahwat) yakni segala yang disenangi serta diingini nafsu sebagai cobaan dari Allah atau tipu daya dari setan (yaitu wanita-wanita, anak-anak dan harta yang banyak) yang berlimpah dan telah berkumpul (berupa emas, perak, kuda-kuda yang tampan) atau baik (binatang ternak) yakni sapi dan kambing (dan sawah ladang) atau tanam-tanaman. (Demikian itu) yakni yang telah disebutkan tadi (merupakan kesenangan hidup dunia) di dunia manusia hidup bersenang-senang dengan hartanya, tetapi kemudian lenyap atau pergi (dan di sisi Allahlah tempat kembali yang baik) yakni surga, sehingga itulah yang seharusnya menjadi idaman dan bukan lainnya.

Ayat ini mengajarkan kita untuk beriman kepada Allah SWT dengan menginfakkan harta kita di jalan kebaikan yang di ridhoi Allah SWT. Yang kita niatkan dengan niat ikhlas, harta kita untuk anak-anak kita belajar di sekolah supaya bisa menjadi ladang pahala menuju surganya.

Oleh karena itu, para orang tua/ wali murid dan pemerintah bertanggung jawab terhadap pendidikan anak-anak, diantaranya adalah perintah untuk membaca dan belajar bagi individu maupun penduduk suatu negara. orang tua wali murid dan pemerintah juga bertanggung jawab terhadap suksesnya kegiatan pendidikan ini. Agar pendidikan berjalan dengan lancar maka perlu disertai dengan biaya pendidikan. Hal tersebut tersirat pada surat Al- Alaq ayat 1-5:

إِفْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ۚ وَإِنَّا لَكَرِيمٌ
الَّذِي عَلَّمَ بِالْقَلَمِ ۗ عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ

Artinya : Dengan menyebut nama Tuhan Penciptamu, bacalah. Dia menjadikan manusia dari segumpal darah. Bacalah dengan teliti, dan Penguasamu adalah Yang Maha Bijaksana, Yang menginstruksikan (manusia) melalui firman, Dia menunjukkan kepada manusia apa yang tidak dia ketahui”.

Sesuai tafsir Al- Jalalain yakni “(Bacalah) maksudnya mulailah membaca dan memulainya (dengan menyebut nama Rabbmu yang menciptakan) semua makhluk. (Dia telah menciptakan manusia) atau jenis manusia (dari 'alaq) lafal 'Alaq bentuk jamak dari lafal 'Alaqah, artinya segumpal darah yang kental. (Bacalah) lafal ayat ini mengukuhkan makna lafal pertama yang sama (dan Rabbmulah Yang Paling Pemurah) artinya tiada seorang pun yang dapat menandingi kemurahan-Nya. Lafal ayat ini sebagai Haal dari Dhamir yang terkandung di dalam lafal Iqra'. (Yang mengajar) manusia menulis (dengan qalam) orang pertama yang menulis dengan memakai qalam atau pena ialah Nabi Idris a.s. (Dia mengajarkan kepada manusia) atau jenis manusia (apa yang tidak diketahuinya) yaitu sebelum Dia mengajarkan kepadanya hidayah, menulis dan berkreasi serta hal-hal lainnya. Pada ayat tersebut mengajak dan memerintahkan belajar, menulis, dan membaca karena ketiganya adalah syiar Islam.

BAB VII

KESIMPULAN DAN SARAN

7.1. Kesimpulan

Berdasarkan analisis dan pembahasan hasil penelitian klasifikasi dalam memprediksi kondisi pembayaran biaya pendidikan sekolah, maka di ambil kesimpulan sebagai berikut:

1. Berdasarkan lima percobaan dengan masing-masing percobaan ada empat skenario, maka hasil penelitian menyatakan bahwa metode Decision Tree layak digunakan untuk memprediksi kondisi pembayaran biaya pendidikan sekolah pada Madrasah Ibtidaiyah Miftahul Huda Sukolilo, dengan hasil akurasi = 71%, kelas tepat presisi = 72%, recall= 99% , dan F1-score= 84%, kelas terlambat presisi = 0%, recall= 0% , dan F1-score= 0%, dan kelas tidak bayar presisi = 0%, recall= 0% , dan F1-score= 0%
2. Akurasi tertinggi pada metode naïve bayes didapatkan dari tiga parameter yaitu (pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan orang tua) yang digunakan, dengan hasil akurasi = 71%, kelas tepat presisi = 72%, recall= 1% , dan F1-score= 84%, kelas terlambat presisi = 0%, recall= 0% , dan F1-score= 0%, dan kelas tidak bayar presisi = 0%, recall= 0% , dan F1-score= 0%, maka dapat disimpulkan faktor- faktor yang berpengaruh dalam mengklasifikasi ekonomi siswa sekolah adalah pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan orang tua. Jadi pekerjaan orang tua sebagai guru, penghasilan cukup dan jumlah tanggungan cukup, memiliki resiko terlambat membayar, dengan rasio kesalahan 1%. Sedangkan akurasi tertinggi pada metode decision tree didapatkan dari dua parameter yaitu penghasilan orang tua dan jumlah tanggungan orang tua, dengan hasil akurasi = 71%, kelas tepat presisi = 72%, recall= 99%, dan F1-score= 84%, kelas terlambat presisi = 0%, recall= 0% , dan F1-score= 0%, dan kelas tidak bayar presisi = 0%, recall= 0% , dan F1-score= 0%, maka disimpulkan faktor- faktor yang berpengaruh dalam mengklasifikasi ekonomi siswa sekolah adalah penghasilan orang tua dan jumlah tanggungan orang tua. Penghasilan

orang tua dengan gaji cukup dan jumlah tanggungan cukup beresiko terlambat dalam membayar biaya pendidikan sekolah, dengan tingkat rasio kesalahan 1.36%.

7.2.Saran

Berdasarkan hasil penelitian, penulis mengemukakan saran untuk menjadi bahan pertimbangan dipenelitian selanjutnya, diantaranya:

1. Penelitian ini hanya menggunakan model prediksi naïve bayes dan decision tree, diharapkan untuk penelitian selanjutnya menambahkan atau menguji model prediksi lain seperti KNN, SVM dan model prediksi lainnya untuk mendapatkan hasil yang lebih baik.
2. Terdapat beberapa variabel dari masing-masing model prediksi yang tidak memiliki pengaruh terhadap kondisi pembayaran biaya pendidikan sekolah. Maka sebaiknya bagi penelitian selanjutnya diperlukan referensi dan pemahaman yang lebih dalam menerapkan model prediksi.
3. Model prediksi selanjutnya diharapkan mampu mencari model prediksi terbaik di berbagai tingkatan sekolah.
4. Memperluas periode waktu sampel agar dapat melihat kondisi pembayaran biaya pendidikan sekolah lebih lengkap.

DAFTAR PUSTAKA

- Abdullah, R. W., Kusriani, & Luthfi, E. T. (2019). Prediksi Keterlambatan Pembayaran Spp Sekolah Dengan Metode K-Nearest Neighbor (Studi Kasus Smk Al-Islam Surakarta). *Jurnal Informasi Interaktif*, 4(3), 1–18.
- Apandi, T. H., Maulana, R. B., Piarna, R., & Vernanda, D. (2019). Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung). *Jurnal Techno Nusa Mandiri*, 16(2), 93–98. <https://doi.org/10.33480/techno.v16i2.659>
- Arhami, M., & Nasir, M. (2020). *DATA MINING ALGORITMA DAN IMPLEMENTASI*. CV ANDI OFFSET.
- Bahtiar, A., & Firmansyah, D. (2017). Klasifikasi Ketepatan Waktu Pembayaran SPP Di Pondok Pesantren Al-Arifah Menggunakan Algoritma Naive Bayes. *Ilmiah Manajemen Informatika Dan Komputer*, XX(Xx), 1–8.
- Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44. <https://doi.org/10.35585/inspir.v10i1.2535>
- Iskandar, I. D. (2019). Parents' Sum of Salaries Analyses towards School Tuition Fee Arrears Potential with Decision Tree Method. *Indonesian Journal of Information Systems*, 2(1), 45–56. <https://doi.org/10.24002/ijis.v2i1.2168>
- Muqorobin, M., Kusriani, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *Jurnal Ilmiah SINUS*, 17(1), 1. <https://doi.org/10.30646/sinus.v17i1.378>
- Muqorobin, M., Kusriani, K., Rokhmah, S., & Muslihah, I. (2020). Estimation System For Late Payment Of School Tuition Fees. *International Journal of Computer and Information System (IJCIS)*, 1(1), 1–6. <https://doi.org/10.29040/ijcis.v1i1.5>
- Prabowo, C., Hakim, L. N., Rohmat, C. L., & Rinaldi, A. R. (2021). Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran. *JURNAL DATA SCIENCE & INFORMATIKA*, 1(1), 1-5., 1(1)(Vol. 1 No. 1 (2021): JDSI-AGUSTUS), 1–5. <http://publikasi.bigdatascience.id/index.php/jdsi/article/view/2>
- Purba, L. P., Hartama, D., Irawan, E., & Wanto, A. (2019). Memprediksi Faktor Tunggakan Uang Kuliah Menggunakan Metode Naive Bayes. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(November), 298.

<https://doi.org/10.30645/senaris.v1i0.35>

Yaqin, F. A., Sumbullah, U., & Rofiq, A. (2021). Manajemen Pembiayaan Pendidikan Ditinjau Tafsir Dan Hadits Tematik. *Journal EVALUASI*, 5(2), 274. <https://doi.org/10.32478/evaluasi.v5i2.741>

LAMPIRAN-LAMPIRAN**Lampiran 1 Tabel Encoder**

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
1	A'an Krono Aji	L	5	1	1	1
2	Aan Rian Pratama	L	2	1	1	2
3	Abdillah Dwi Kurnia Santoso	L	1	0	1	1
4	Abellita Yuniar Syaputri	L	3	3	1	1
5	Achmad Arifin	L	2	3	1	0
6	Achmad Fawaidul Ilmi	L	0	0	1	1
7	Achmad Galih Permana	L	2	1	1	1
8	Achmad Revi Firmansyah	P	4	1	1	0
9	Addini Nafisatur Rohmah	P	4	1	1	2
10	Adelia Zahira	P	3	1	0	1
11	Adinda Qhurotul Khasanah	P	2	1	0	1
12	Adinda Sasa Ayu	P	5	3	1	1
13	Afrizal Dhifan Rizki Septian	P	2	1	0	1
14	Afuwwu	P	2	3	0	2
15	Agam Bagus Julianysah	L	2	3	1	1
16	Aghis Sabillah Syafa Aushony	P	4	3	1	1
17	Aghna Zida Muhammad	L	2	1	1	1
18	Agung Habibi	P	3	1	0	1
19	Agustama Zaidan Maliki	L	4	3	0	1
20	Ahdan Nurillah Tsaqif	L	2	1	1	1
21	Ahmad Ali Furqon Romadhoni	L	0	0	0	0
22	Ahmad Bahril Ilmi Daviq	L	4	0	1	1
23	Ahmad Bilal Mauludin	L	2	0	1	0
24	Ahmad Daffa Hafidul Ahkam	L	3	3	1	1
25	Ahmad Ihsamudin Burhan	L	1	2	1	0
26	Ahmad Ilham Syahril Muafiq	L	2	3	0	1
27	Ahmad Khunaifi	L	3	1	1	2
28	Ahmad Nabil Khasbillah Rojabi	L	3	3	1	1
29	Ahmad Revalino Dwi Syaputra	L	4	3	1	2
30	Ahmad Reza Mustofa	L	1	1	1	2
31	Ahmad Rifky Ariansyah	L	2	1	0	0
32	Ahmad Rizky Adiharja	L	2	3	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
33	Ahmad Syarif Hidayatulloh	L	3	3	0	1
34	Ahmad Yasrif Fachry Muhajir	L	5	3	1	0
35	Aini Maisya Putri Cantika	P	2	3	1	1
36	Ainun Zahkia Niji	P	4	3	1	1
37	Aisyah Aghna Syahira	P	4	3	1	1
38	Akroma Wafiroh	P	2	3	1	0
39	Al - Mar'atus Syarifah	P	3	3	1	1
40	Al Ashfahani Khuurun Azzahra	P	2	1	0	1
41	Aldhanea Mifta Saffela	P	2	1	0	1
42	Aldiano Bimo Pratama	L	4	1	0	1
43	Alfan Ilham Ramadhani	L	2	1	1	1
44	Alfina Rizqia Kamila	P	0	1	1	1
45	Alfira Ria Agustina	P	2	0	1	1
46	Aliffatul Nur Jannah	P	2	0	1	1
47	Alifia Nada Ramadhani	P	4	3	1	1
48	Alin Fembi Anjani Stevani	P	5	0	1	1
49	Alisha Khaira Zevanna	P	1	0	1	1
50	Alldi Eka Saputra	P	2	2	0	2
51	Almira Najwa Monieza	P	2	1	1	2
52	Alun Dwi Hamdansyah	L	3	3	0	0
53	Alzahira Dewi Elzaki Ramadhani	P	2	0	1	2
54	Amar Haidar Alzam	L	2	1	1	1
55	Amelia Nuril Khomariya	P	2	1	1	0
56	Amelia Rahma Hapsari	L	5	0	1	1
57	Amin Rakha Ramadhan	P	0	1	1	0
58	Ana Afidatul Khuzna	P	2	3	1	1
59	Anfasa Aisyl Alan	L	2	1	0	1
60	Angga Dwi Pratama	P	4	3	1	0
61	Anggun Dian Putri Lestari	L	4	0	1	2
62	Aninda Selly Maulida	L	5	1	0	1
63	Anisa Permata Indah Safitri	P	4	1	1	1
64	Anom Baihaqi	P	4	0	1	0
65	Anom Muzaky	P	1	3	1	1
66	Aqilla Azahwa Algivary	L	2	3	1	2
67	Ardhin Bila Harbi El Muslim	L	3	0	1	2

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
68	Arfan Maulana Syahputra	P	2	1	1	1
69	Aria Putra Pratama	L	2	1	0	1
70	Arin Dina Islami	L	4	1	1	0
71	Arini Amalia	L	1	1	1	1
72	Aris Total Onasis	P	3	1	1	1
73	Arsa Adibrata	P	2	3	1	1
74	Arvelino Yafattah Primadasena	L	0	1	1	1
75	Arya Rizki Agung Al Hidayat	L	1	2	0	1
76	Arzeta Carissa Early	L	2	3	1	1
77	Asyam Farkhan Raif	L	5	3	1	1
78	Asyifah Auliatur Rahmah	P	5	1	0	1
79	Atha Aditya Zaidan	L	2	1	1	1
80	Atha'illah Wahidin Rahman Nafis	P	2	3	1	1
81	Athaya Ramadhani Whildan Firdausi	L	2	1	0	1
82	Athiyyah Nur Aini	L	1	0	0	1
83	Audrea Danurdara Aglaeca	L	4	0	1	1
84	Aulya Wahyuningtyas	L	0	0	1	0
85	Azam Adzzikriansyah	P	2	3	1	0
86	Azizah Rahma Fariesya	P	2	3	1	1
87	Azka Alif Pratama	P	4	3	1	0
88	Azka Dzikro Sofiya Fahmi	P	2	1	0	2
89	Azka Tsania Az-Zahra	L	2	3	0	1
90	Azmil Mufida	P	3	3	1	0
91	Azril Faris Putra Laksmana	L	2	1	1	1
92	Bimas Darma Rubiyanto	L	5	1	1	1
93	Bisma Pratama	P	2	3	1	1
94	Carissa Fidela Susena	P	2	3	1	1
95	Carissa Qurrotayun Azzahra	L	2	3	1	1
96	Cellova Alingga Navy Armadyra	L	2	3	1	0
97	Chairus Syuhud	L	4	3	1	1
98	Chesta Harina Adabi	L	4	3	1	1
99	Cintami Atmanegara Artanti	P	4	3	1	1
100	Daffa Fauzi Nur Yatim	L	2	3	0	1
101	Damai Permata Efendy	L	4	1	0	0

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
102	Darin Al Atira	P	1	1	0	2
103	Dava Ainu Vikry	L	5	1	1	1
104	Dava Marvel Saputra	L	2	1	1	1
105	Dermawan Adi Prakoso	L	1	1	1	1
106	Devinno Adrian Putra Maulana	L	2	0	1	0
107	Dewi Fathonah	P	2	0	0	1
108	Diah Ayu Puspita	P	2	3	1	1
109	Diandra Marsudin	L	2	0	0	1
110	Dimas Dwi Ari Saputra	L	1	0	1	1
111	Dina Putri Arulia	P	2	1	1	2
112	Dinda Fadhia Maulidia	P	2	1	1	1
113	Diva Athira Rahmah	P	1	3	0	1
114	Dwi Nursusanti	P	3	0	1	1
115	Ela Putri Damayanti	P	3	1	1	1
116	Elang Goldy Mahendra	L	1	1	1	0
117	Eliana Ramadhani	P	2	0	1	1
118	Erlangga Dwi Alvino	L	3	1	1	1
119	Esa Alfiansyah	L	5	3	0	1
120	Eva Mardiana	P	2	1	1	1
121	Fahmy Adi Amrulloh	L	1	3	1	1
122	Fahri Adis Maulana	P	3	0	1	1
123	Fahri Cahaya	L	4	1	1	0
124	Faldan Okta Andira	L	4	1	1	2
125	Fania Cintana Putri Anggraini	P	5	0	1	0
126	Fania Hikmah Ulia	P	5	3	1	2
127	Farah Daimatu Nafisa	P	0	3	1	1
128	Farhan Assidiqi	L	3	0	1	1
129	Farisha Shanum Hardiansa	P	3	1	1	0
130	Fatma Qurrotuaini	P	2	1	1	1
131	Fawatikha Riska Fauziyah	P	2	1	1	1
132	Ferdio Kurnia Putra	L	2	1	0	1
133	Ferreira Ashley Dhe Evelyn	P	4	1	0	1
134	Fiko Bayu Armada	L	1	3	1	1
135	Fira Zafiliya	P	3	1	1	0
136	Firmi Augista Ramadhan	P	2	3	1	1
137	Fitah Ainis Qornaini	P	4	3	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
138	Fitrotul Mufida	P	4	3	1	1
139	Galih Dwi Prasetyo	L	2	1	1	1
140	Geisa Dewi Liawati	P	1	1	1	1
141	Gelsi Hilwana Azalea	P	2	3	1	1
142	Habibie Dryan Hamungkasih	L	1	1	0	1
143	Hafiz Rakha Nadhif	L	3	0	0	1
144	Hafizah Khanza Azzahra	P	0	0	1	1
145	Haira Ibriza Rayani	P	2	0	1	1
146	Handinata Prastio	L	5	3	1	1
147	Hari Mulyono	L	4	3	1	1
148	Hasna Laily Syarifah Hanin Siregar	P	2	3	0	1
149	Haura Damia Azzahra	P	2	1	1	1
150	Hilda Salwa Fitria	P	3	3	0	2
151	Humam Dafa Muhammad	L	3	3	1	1
152	Icha Ananda Putri Vela Tyas Wati	P	2	1	1	1
153	Ihda Annisa Khumairo	P	2	1	1	2
154	Ilham Achmad	L	5	3	1	0
155	Imam Ghozali	L	1	3	1	1
156	Intan Naila Kartika Sari Devi	P	1	3	1	1
157	Irham Wildani Robby	L	4	3	1	1
158	Isma Az Zahra Salsabilla	P	2	3	1	2
159	Ityanu Al Fauzan	L	2	3	1	1
160	Ivane Naila Yuna	P	1	3	1	1
161	Izzati Zakiatus Syarifah	P	3	3	1	1
162	Jazilatul Husniyah Alkarima	P	2	1	1	1
163	Jihan Inggrit Arizta	P	4	1	1	1
164	Jihan Putri Aprilia	P	3	1	0	1
165	Junita Pandu Winata S.	P	4	1	1	1
166	Kamila Afkarima	P	2	1	1	2
167	Kamilah Ainun Nisa'	P	4	0	0	1
168	Kayla Nahda Bilqis	P	2	0	1	1
169	Keane Dharma Putra Alfarezel	P	5	3	1	1
170	Keisya Beby Octavia	P	5	0	1	0
171	Kenzie Sadina Aryasatya	L	2	0	1	1
172	Keyla Putri Anggraini	P	2	1	0	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
173	Keysyafa Aqiela Elqisyah Arifianty	P	2	1	0	1
174	Khaira Aulia Azzahra	P	3	3	0	0
175	Khanza Dhiyaul Huwaidah Widiyanti	P	2	0	1	1
176	Khanza Nailah Susena	P	4	1	1	1
177	Khanza Neysa Dwi Putri Anggraeni	P	3	1	0	1
178	Khofidah Nur Maulidiyah	P	2	0	1	1
179	Khoirun Nisa'	P	4	1	1	1
180	Kirein Hanania Salsabila	P	0	3	1	1
181	Kireina Nawang Nurinnajwa	P	4	1	0	1
182	Kurniawan Adi Al - Kautsar	L	2	3	1	1
183	Laura Ismaya Bilqis	P	2	0	0	1
184	M Kharis Maulidan	L	5	1	1	1
185	M Naufal Azmi Alauddin	L	2	1	1	1
186	M Raihan Nasrullah Shidqi	L	2	0	1	1
187	M Umar Al Hafidz	L	2	3	0	1
188	M. Alfa Fikri Maulana	L	4	3	0	0
189	M. Alfin Fikri Azzahron	L	1	0	0	1
190	M. Alif Pradana	L	4	1	1	1
191	M. Alvin Atian Syabana	L	5	1	1	1
192	M. Asyrafil Khalqi Huda	L	2	1	1	1
193	M. Bachrul Ilmi Romadhon	L	1	1	1	2
194	M. Bagus Bening Wicaksono	L	2	1	1	1
195	M. Barik Fauzan Adzim	L	3	3	1	1
196	M. Daroini Faizin	L	1	1	0	0
197	M. Denish Al-Baihaqi	L	3	3	1	1
198	M. Faturrahmah Hakim	L	2	3	1	1
199	M. Fazriel Dwi Irzami	L	1	3	1	0
200	M. Fikri Haikal	L	5	1	1	1
201	M. Ilham Rafi Wildana	L	4	1	1	1
202	M. Labib Amrulloh Abdalah	L	2	3	0	1
203	M. Misbahur Ridho	L	2	1	0	1
204	M. Nizam Nur Rohim	L	0	0	1	1
205	M. Pandu Indra Prasetyo	L	2	0	1	0
206	M. Ridlwan A.R	L	5	0	1	1
207	M. Rijal Fadlli	L	3	3	0	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
208	M. Roihan Azzuhry Zamakhsyari	L	0	3	1	1
209	M. Romi Ansorullah	L	0	3	1	1
210	M. S Khoirul Fahmi	L	2	1	0	1
211	M. Syahrul Fadli Maulana	L	2	3	1	2
212	M. Syaiful Anwar Riyadi	L	4	3	1	0
213	M. Syaroyyan Tsahwi El Wafi	L	2	1	0	1
214	M. Taufiq Hidayatulloh	L	2	1	1	1
215	M. Urfiyul Fahmi Maulana	L	2	3	1	1
216	M. Zainul Furkon Ramadani	L	4	3	1	0
217	M. Zidni Hayatal Mubarak	L	2	3	0	1
218	Mahira Hasna Astila Rahma	P	2	3	0	1
219	Mahira Salma Qotrun Nada	P	2	3	1	1
220	Maulana Akbar Almubarak	L	4	3	1	0
221	Maulana Iqbal Bawazir	L	5	3	1	1
222	Maulana Muhammad Fachri	L	3	3	1	1
223	Maulidul Rizqil Kubro	L	2	1	0	0
224	Mazaya Salsabila	P	1	1	1	1
225	Meicha Anggraeni	P	5	1	1	0
226	Melinda Fitrianti	P	4	1	0	1
227	Metyu Stiven Rahmad Dani	L	2	1	1	1
228	Milasofi Meida	P	2	0	1	1
229	Mira Mustika	P	1	0	1	2
230	Mirza Nery Hatka Rajasa	P	2	3	0	2
231	Mizan Ulil Alwan	L	1	0	0	1
232	Mochamad Rizky Ramadhani	L	5	0	1	2
233	Mochammad Ircham Milady	L	3	1	0	1
234	Mochammad Ridho Alamsyah	L	4	1	0	1
235	Mochammad Sandi Prayoga	L	2	3	1	1
236	Moh. Maulana Riska	L	2	0	1	1
237	Mohammad Adrian Maulana	L	5	1	1	1
238	Muchamad Abdul Faiz	L	2	1	0	1
239	Muchammad Raditya Maulana Rizqi	L	5	0	0	1
240	Mudjiyati	P	1	1	0	1
241	Mufidah Zahroh	P	2	3	1	2
242	Mufidatun Najwa Salsabila	P	3	1	0	0

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
243	Muh. Danu Wanda Pratama	L	3	3	0	1
244	Muh. Dicky Pratama	L	2	0	1	2
245	Muhamad Aditya Eka Putra	L	2	1	1	1
246	Muhamad Falevi Romadhon	L	4	1	1	1
247	Muhamad Faza Adzima Syauqi Al Ghifari	L	4	0	1	2
248	Muhamad Tegar Khaidar Wibawanto	L	2	3	0	1
249	Muhammad Abdillah Iman Pangestu	L	2	3	1	1
250	Muhammad Abid Tamami	L	4	0	1	1
251	Muhammad Aby Satrio Lastavano	L	5	1	0	1
252	Muhammad Adam Dwi Andhika	L	2	1	1	1
253	Muhammad Adit Tiansyah	L	3	1	1	1
254	Muhammad Aditya Adi Darma	L	4	1	1	1
255	Muhammad Aditya Candra Pratama	L	4	1	1	1
256	Muhammad Aditya Firmansyah	L	3	3	1	0
257	Muhammad Aisaru Wildan	L	2	1	1	2
258	Muhammad Akbar Rangga Dinata	L	2	3	0	1
259	Muhammad Aldo Setyawan	L	1	3	1	1
260	Muhammad Alfarizi	L	2	3	1	0
261	Muhammad Alfin Zulkifli	L	2	1	0	1
262	Muhammad Ali Fauzan	L	1	1	0	1
263	Muhammad Ali Wafa	L	2	3	1	1
264	Muhammad Alvaro Ahnaf Afandi	L	2	1	1	1
265	Muhammad Amin	L	0	0	1	2
266	Muhammad Andi Prayoga	L	3	0	1	1
267	Muhammad Andika Saputra	L	2	0	0	1
268	Muhammad Aqil Dzikry	L	4	3	1	1
269	Muhammad Arvin Daniel Alfaro	L	2	3	1	1
270	Muhammad Averroes Ali Sya'bana	L	5	3	0	1
271	Muhammad Awangga Pratama	L	3	1	1	0
272	Muhammad Azriel Al Hafiidl	L	1	3	1	1
273	Muhammad Baihaqi Khaizan	L	2	3	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
	Alifuddin					
274	Muhammad Barqi Mahasin	L	2	1	1	0
275	Muhammad Dinar Al Fairus	L	1	1	1	1
276	Muhammad Eka Sya'bana	L	4	3	1	1
277	Muhammad Fahmi Abimanyu	L	2	3	1	1
278	Muhammad Fahmi Zidny	L	2	3	1	2
279	Muhammad Farhan Azzam Pratama	L	0	3	0	1
280	Muhammad Fikri Ardiansyah	L	2	3	1	1
281	Muhammad Gilang Ramadhan	L	2	3	1	1
282	Muhammad Gilbran	L	3	3	1	0
283	Muhammad Ighfirlana Ramadani	L	2	3	1	0
284	Muhammad Iman Nur Diansyah	L	1	1	1	1
285	Muhammad Iqbal Maulana	L	1	1	0	1
286	Muhammad Irfan Al Arsyad	L	4	1	1	0
287	Muhammad Irham Bighvirly	L	3	1	0	2
288	Muhammad Irham Pratama Nurrohmad	L	2	1	1	2
289	Muhammad Isya Aidil Ahkam	L	2	0	1	0
290	Muhammad Jibril Al-Fariz	L	5	0	1	1
291	Muhammad Khasanil Khuluk	L	0	3	1	0
292	Muhammad Khoirul Anam	L	1	0	1	1
293	Muhammad Kholili	L	3	0	1	1
294	Muhammad Luqman Al Hakim	L	2	1	1	1
295	Muhammad Muhtarul Adzim	L	2	1	1	2
296	Muhammad Muslimin	L	2	3	1	1
297	Muhammad Nafis Sakha Anwar	L	2	0	0	1
298	Muhammad Naufal	L	3	1	1	1
299	Muhammad Naufal Asyrafu Muzakki	L	2	1	1	1
300	Muhammad Naufal Brilian Azzam Envayya	L	2	0	1	1
301	Muhammad Nur Ikhsan	L	2	1	1	1
302	Muhammad Nur Izam Maulana	L	2	3	1	1
303	Muhammad Ridho Rofi Qunazili	L	2	1	0	0
304	Muhammad Rifqi Fairuz	L	3	3	1	2

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
305	Muhammad Rizky Irsyad Latief	L	2	0	1	1
306	Muhammad Rizky Pramudika	L	0	1	0	1
307	Muhammad Royyan Arrohman	L	2	1	1	1
308	Muhammad Sifaul Qulub	L	4	0	1	1
309	Muhammad Syauqi	L	2	3	1	1
310	Muhammad Taqiyuddin Al Ghozah	L	2	3	1	1
311	Mukhamat Ahsanul Kholikin	L	3	0	0	1
312	Mukhamat Iqvurul Aziz	L	3	1	0	0
313	Mukhammad Aris Ar Ridho	L	2	1	0	1
314	Mukhammad Aris Firdianto	L	5	1	1	1
315	Mukhammad Rizqi Saputra	L	3	1	0	1
316	Nabila Sandra Aurelya Ramadhani	P	5	1	0	1
317	Nabila Saskia Putri Islami	P	2	3	1	1
318	Nabrisatul Arifah	P	2	1	0	1
319	Nadia Ahla Salsabila	P	4	3	1	1
320	Nadia Hafidzah Al Wahda	P	1	3	1	0
321	Nafasad Nizam Istibra	L	4	3	1	1
322	Nafisa Khayla Amanda	P	4	1	1	1
323	Nafisha Ayda Ramadhani	P	2	1	1	2
324	Naila Istiadzah	P	2	3	1	1
325	Naila Khafiyah	P	3	1	1	1
326	Naisyifa Putri Almira	P	2	0	1	0
327	Najma Zahira	P	2	0	1	0
328	Najmatul Ilmi	P	5	0	0	2
329	Najwa Faradissa	P	2	3	1	1
330	Najwa Nafisa Adabina	P	4	3	1	1
331	Nanda Kurniawati	P	3	3	1	1
332	Nasya Akila Zifana	P	2	1	1	1
333	Nathana Aufa Azzahra	P	2	3	1	0
334	Naura Hasna Aninda	P	2	3	1	1
335	Navella Damai Yanti Mustariza	P	4	1	1	1
336	Nayla Fariza Azmi	P	5	1	1	2
337	Nico Bari' Pratama	L	2	3	1	1
338	Niken Adinda Pratiwi	P	2	3	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
339	Nimas Ayu Ramadani	P	2	3	1	0
340	Nizam Ardian Permana	P	2	3	1	0
341	Noval Aditya Rifaldi	L	5	3	1	0
342	Novelli Dwi Yanti Mustariza	P	1	3	1	1
343	Novi Sekar Ayu	P	3	3	1	1
344	Nur Afifah Aulia	P	4	3	1	1
345	Nur Alifatul Khuriyah	P	2	1	0	1
346	Nur Niesfatul Fitria Romadhoni	P	3	1	1	1
347	Nur Olivia	P	2	1	1	1
348	Nur Wahyu Ningsih	P	2	1	1	1
349	Nur Zahra An Nabila	P	2	1	1	1
350	Nurdy Handayani	P	2	0	0	0
351	Nyla Azzahra Asylla Rahma	P	2	0	1	0
352	Panji Nugroho Saputra	L	5	3	0	0
353	Prisa Adinda Alfarini Eridani	P	2	0	1	1
354	Priska Eka Putri	P	2	0	1	1
355	Putra Rizki Firmansyah	L	2	1	1	1
356	Putri Tsania Nawang Ayuprarnesti	P	5	1	1	1
357	Qonita Isman Taqwiya	P	2	3	1	1
358	Queen Alhysia Erly Firmansyah	P	3	0	1	0
359	Queena Artantia Putri Hartono	P	5	1	1	1
360	Rafa Fatina Putri	P	2	1	1	2
361	Rafa Martiza	P	2	0	1	0
362	Rahmadina Nuraini	P	2	1	1	1
363	Rahmat Fitra Zakaria	L	5	3	1	1
364	Raihan Ramadhani Alfatih	L	5	1	1	1
365	Raihana Aura Putri	P	4	3	1	1
366	Raka Al-Varo Syahri Al- Pratama	L	2	0	1	1
367	Raka Ananda Pratama	L	2	1	1	1
368	Ra'kha Arda Sugiarto	L	2	1	1	1
369	Rangga Nor Setiawan	L	2	0	1	1
370	Rendy Septian Nugroho	L	1	3	1	1
371	Riska Andini	P	2	3	1	1
372	Risky Aisi Putra	L	2	0	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
373	Rista Aprilia	P	1	1	1	1
374	Riyo Prawira Saputra	L	2	1	0	1
375	Rizky Romadhoni	L	4	1	0	0
376	Rizqi Rahmad Ramadhani	L	2	1	1	1
377	Rizqy Radita Pratama	L	2	1	1	1
378	Rofiyatul Qoyimah	P	0	3	1	1
379	Rosdiana Ayu Riski	P	2	1	1	1
380	Roudlotul Janah	P	2	3	1	2
381	Safara Andinia Putri	P	2	3	1	1
382	Sanyatu Alifa Aprilia	P	2	3	0	0
383	Satria Anggoro	L	4	1	1	0
384	Satria Pratama Mahardika	L	0	1	0	2
385	Sayyidah Yasmin Al Maghfira	P	0	3	1	2
386	Shania Kotrun Nada	P	3	1	1	1
387	Shasmecka Aira Nugrahanti	P	2	0	1	0
388	Sheila Aprilia Haqiqi	P	5	0	1	1
389	Sifa Ayu Kinanda	P	0	0	1	0
390	Siti Maisaroh	P	2	3	1	1
391	Siti Mariyam Wulandari	P	2	3	1	2
392	Siti Nur Laili Khumairotu Zahra	P	2	3	0	1
393	Slamet Wildan Misbakhul Akfa	L	4	1	1	1
394	Sri Mulyarsih	P	2	3	1	0
395	Suci Amalia	P	2	3	0	1
396	Sultan Yafi Saputro	L	2	1	1	1
397	Syafika Aulia	P	2	1	1	1
398	Syakira Jovita	P	2	3	1	1
399	Syifaun Nasya Alifa	P	3	3	0	2
400	Trisna Al Kahfi Bintara	L	5	3	1	0
401	Tsamrotul Hidayah	P	1	3	1	1
402	Uliya Tahfidhotul Karomah	P	3	3	0	2
403	Umi Hanik	P	2	3	1	1
404	Umi Nafilla	P	2	3	0	1
405	Unwanun Nadhif Maulidil Ghofiri	L	2	3	1	1
406	Vanessa Olivia Aryani	P	3	1	1	1
407	Vrika Early Maulida	P	4	1	1	1

No	Nama	Kelamin	Pekerjaan _Orang_ Tua	Penghasilan	Jumlah_Tang gungan	Actual_ Class
408	Wahyu Lisa Muthoharoh	P	3	1	1	1
409	Widiyanti Fitri Ramadhan	P	4	1	0	1
410	Zabdan Ramadhan Qurota Ayun	L	2	1	0	1
411	Zacky Irdian Sutrisno	L	2	0	1	1
412	Zahira Qothrun Nada	P	4	0	1	1
413	Zahira Rihadatul Aisi	P	3	3	1	1
414	Zahra Afista Nanda	P	1	0	1	2
415	Zahrhatus Sytha	P	4	0	0	1
416	Zahrotul Fitriyatus Sholichah	P	0	1	1	0
417	Zahrotul Jannah	P	5	1	1	2
418	Zahrotul Sarirotul Mukaromah	P	4	3	0	0
419	Zaki Ahmad Ramadhani	L	1	0	1	1
420	Zaky Syahbana Putra Wicaksono	L	2	1	1	0

Lampiran 2 Tabel komparasi naïve bayes dan decision tree

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
1	M. Syahrul Fadli Maulana	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
2	M. Syaiful Anwar Riyadi	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
3	M. Syaroyyan Tsahwi El Wafi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
4	M. Taufiq Hidayatulloh	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
5	M. Urfiyul Fahmi Maulana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
6	M. Zainul Furkon Ramadani	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
7	M. Zidni Hayatal Mubarak	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
8	Mahira Hasna Astila Rahma	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
9	Mahira Salma Qotrun Nada	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
10	Maulana Akbar Almubarak	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
11	Maulana Iqbal Bawazir	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
12	Maulana Muhammad Fachri	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
13	Maulidul Rizqil Kubro	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
14	Mazaya Salsabila	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
15	Meicha Anggraeni	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
16	Melinda Fitrianti	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
17	Metyu Stiven Rahmad Dani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
18	Milasofi Meida	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
19	Mira Mustika	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
20	Mirza Nery Hatka Rajasa	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
21	Mizan Ulil Alwan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
22	Mochamad Rizky Ramadhani	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
23	Mochammad Ircham Milady	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
24	Mochammad Ridho Alamsyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
25	Mochammad Sandi Prayoga	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0

No	Nama	Actual Class	Naïve Bayes					Decision Tree				
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
26	Moh. Maulana Riska	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
27	Mohammad Adrian Maulana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
28	Muchamad Abdul Faiz	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
29	Muchammad Raditya Maulana Rizqi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
30	Mudjiyati	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
31	Mufidah Zahroh	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
32	Mufidatun Najwa Salsabila	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
33	Muh. Danu Wanda Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
34	Muh. Dicky Pratama	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
35	Muhamad Aditya Eka Putra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
36	Muhamad Falevi Romadhon	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
37	Muhamad Faza Adzima Syauqi Al Ghifari	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
38	Muhamad Tegar Khaidar Wibawanto	Terlambat	Terlambat		1	2	0	Terlambat	1	2	0	0
39	Muhammad Abdillah Iman Pangestu	Terlambat	Terlambat		1	2	0	Terlambat	1	2	0	0
40	Muhammad Abid Tamami	Terlambat	Terlambat		1	2	0	Terlambat	1	2	0	0
41	Muhammad Aby Satryo Lastavano	Terlambat	Terlambat		1	2	0	Terlambat	1	2	0	0
42	Muhammad Adam Dwi Andhika	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
43	Muhammad Adit Tiansyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
44	Muhammad Aditya Adi Darma	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
45	Muhammad Aditya Candra Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
46	Muhammad Aditya Firmansyah	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
47	Muhammad	Tidak	Terlambat	0	1	1	1	Terlambat	0	1	1	1

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
	Aisaru Wildan	Bayar										
48	Muhammad Akbar Rangga Dinata	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
49	Muhammad Aldo Setyawan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
50	Muhammad Alfarizi	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
51	Muhammad Alfin Zulkifli	Terlambat	Terlambat	1	2	0	0	Tepat	0	1	1	1
52	Muhammad Ali Fauzan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
53	Muhammad Ali Wafa	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
54	Muhammad Alvaro Ahnaf Afandi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
55	Muhammad Amin	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
56	Muhammad Andi Prayoga	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
57	Muhammad Andika Saputra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
58	Muhammad Aqil Dzikry	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
59	Muhammad Arvin Daniel Alfaro	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
60	Muhammad Averroes Ali Sya'bana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
61	Muhammad Awangga Pratama	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
62	Muhammad Azriel Al Hafidil	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
63	Muhammad Baihaqi Khaizan Alifuddin	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
64	Muhammad Barqi Mahasin	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
65	Muhammad Dinar Al Fairus	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
66	Muhammad Eka Sya'bana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
67	Muhammad Fahmi Abimanyu	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
68	Muhammad Fahmi Zidny	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
69	Muhammad Farhan Azzam Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0

No	Nama	Actual Class	Naïve Bayes					Decision Tree				
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
70	Muhammad Fikri Ardiansyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
71	Muhammad Gilang Ramadhan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
72	Muhammad Gilbran	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
73	Muhammad Ighfirlana Ramadani	Tepat	Terlambat	0	1	1	1	Tepat	0	1	1	1
74	Muhammad Iman Nur Diansyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
75	Muhammad Iqbal Maulana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
76	Muhammad Irfan Al Arsyad	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
77	Muhammad Irham Bighvirly	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
78	Muhammad Irham Pratama Nurrohmad	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
79	Muhammad Isya Aidil Ahkam	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
80	Muhammad Jibril Al-Fariz	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
81	Muhammad Khasanil Khuluk	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
82	Muhammad Khoirul Anam	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
83	Muhammad Kholili	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
84	Muhammad Luqman Al Hakim	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
85	Muhammad Muhtarul Adzim	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
86	Muhammad Muslimin	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
87	Muhammad Nafis Sakha Anwar	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
88	Muhammad Naufal	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
89	Muhammad Naufal Asyraf Muzakki	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
90	Muhammad Naufal Brilian Azzam Envayya	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
91	Muhammad Nur Ikhsan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
92	Muhammad Nur Izam Maulana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
93	Muhammad Ridho Rofi Qunazili	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
94	Muhammad Rifqi Fairuz	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
95	Muhammad Rizky Irsyad Latief	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
96	Muhammad Rizky Pramudika	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
97	Muhammad Royyan Arrohman	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
98	Muhammad Sifaul Qulub	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
99	Muhammad Syauiqi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
100	Muhammad Taqiyyuddin Al Ghozah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
101	Mukhamat Ahsanul Kholikin	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
102	Mukhamat Iqvurul Aziz	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
103	Mukhammad Aris Ar Ridho	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
104	Mukhammad Aris Firdianto	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
105	Mukhammad Rizqi Saputra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
106	Nabila Sandra Aurelya Ramadhani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
107	Nabila Saskia Putri Islami	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
108	Nabrisatul Arifah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
109	Nadia Ahla Salsabila	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
110	Nadia Hafidzah Al Wahda	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
111	Nafasad Nizam Istibra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
112	Nafisa Khayla Amanda	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
113	Nafisha Ayda Ramadhani	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
114	Naila Istiadzah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
115	Naila Khafiyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
116	Naisyifa Putri	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
	Almira											
117	Najma Zahira	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
118	Najmatul Ilmi	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
119	Najwa Faradissa	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
120	Najwa Nafisa Adabina	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
121	Nanda Kurniawati	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
122	Nasya Akila Zifana	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
123	Nathana Aufa Azzahra	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
124	Naura Hasna Aninda	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
125	Navella Damai Yanti Mustariza	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
126	Nayla Fariza Azmi	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
127	Nico Bari' Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
128	Niken Adinda Pratiwi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
129	Nimas Ayu Ramadani	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
130	Nizam Ardian Permana	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
131	Noval Aditya Rifaldi	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
132	Novelli Dwi Yanti Mustariza	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
133	Novi Sekar Ayu	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
134	Nur Afifah Aulia	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
135	Nur Alifatul Khuriyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
136	Nur Niesfatul Fitria Romadhoni	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
137	Nur Olivia	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
138	Nur Wahyu Ningsih	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
139	Nur Zahra An Nabila	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
140	Nurdya Handayani	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
141	Nyla Azzahra Asylla Rahma	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
142	Panji Nugroho Saputra	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
143	Prisa Adinda Alfarini Eridani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
144	Priska Eka Putri	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
145	Putra Rizki Firmansyah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
146	Putri Tsania Nawang Ayupramesti	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
147	Qonita Isman Taqwiya	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
148	Queen Alhysia Erly Firmansyah	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
149	Queena Artantia Putri Hartono	Terlambat	Terlambat	1	2	0	0	Terlambat	0	1	1	1
150	Rafa Fatina Putri	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
151	Rafa Martiza	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
152	Rahmadina Nuraini	Terlambat	Terlambat	0	1	1	1	Terlambat	1	2	0	0
153	Rahmat Fitra Zakaria	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
154	Raihan Ramadhani Alfatih	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
155	Raihana Aura Putri	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
156	Raka Al-Varo Syahri Al-Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
157	Raka Ananda Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
158	Ra'kha Arda Sugiarto	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
159	Rangga Nor Setiawan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
160	Rendy Septian Nugroho	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
161	Riska Andini	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
162	Risky Aisi Putra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
163	Rista Aprilia	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
164	Riyo Prawira Saputra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
165	Rizky Romadhoni	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
166	Rizqi Rahmad Ramadhani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
167	Rizqy Radita Pratama	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
168	Rofiyatul	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0

No	Nama	Actual Class	Naïve Bayes				Decision Tree					
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
	Qoyimah											
169	Rosdiana Ayu Riski	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
170	Roudlotul Janah	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
171	Safara Andinia Putri	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
172	Sanyatu Alifa Aprilia	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
173	Satria Anggoro	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
174	Satria Pratama Mahardika	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
175	Sayyidah Yasmin Al Maghfira	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
176	Shania Kotrun Nada	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
177	Shasmecka Aira Nugrahanti	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
178	Sheila Aprilia Haqiqi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
179	Sifa Ayu Kinanda	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
180	Siti Maisaroh	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
181	Siti Mariyam Wulandari	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
182	Siti Nur Laili Khumairotu Zahra	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
183	Slamet Wildan Misbakhul Akfa	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
184	Sri Mulyarsih	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
185	Suci Amalia	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
186	Sultan Yafi Saputro	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
187	Syafika Aulia	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
188	Syakira Jovita	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
189	Syifaun Nasya Alifa	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
190	Trisna Al Kahfi Bintara	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
191	Tsamrotul Hidayah	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
192	Uliya Tahfidhotul Karomah	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
193	Umi Hanik	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
194	Umi Nafilla	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
195	Unwanun Nadhif Maulidil Ghofiri	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0

No	Nama	Actual Class	Naïve Bayes					Decision Tree				
			Hasil Prediksi	TP	TN	FP	FN	Hasil Prediksi	TP	TN	FP	FN
196	Vanessa Olivia Aryani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
197	Vrika Early Maulida	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
198	Wahyu Lisa Muthoharoh	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
199	Widiyanti Fitri Rahmadan	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
200	Zabdan Rahmadan Qurota Ayun	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
201	Zacky Irdian Sutrisno	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
202	Zahira Qothrun Nada	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
203	Zahira Rihadatul Aisi	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
204	Zahra Afista Nanda	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
205	Zahrhatus Sytha	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
206	Zahrotul Fitriyatus Sholichah	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
207	Zahrotul Jannah	Tidak Bayar	Terlambat	0	1	1	1	Terlambat	0	1	1	1
208	Zahrotul Sarirotul Mukaromah	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1
209	Zaki Ahmad Ramadhani	Terlambat	Terlambat	1	2	0	0	Terlambat	1	2	0	0
210	Zaky Syahbana Putra Wicaksono	Tepat	Terlambat	0	1	1	1	Terlambat	0	1	1	1

Lampiran 3 Pseudocode

```
import numpy as np
import pandas as pd
import time
from google.colab import drive
drive.mount('/content/gdrive', force_remount=True)
dataset = pd.read_csv("gdrive/MyDrive/Colab Notebooks/TESIS2/d
ata2.csv")
dataset
def drop_data():
    print ('Proses drop kolom yang tidak diperlukan')
    global dataset
    dataset = dataset.drop(columns=['NO', 'NAMA', 'KELAMIN', 'JUM
LAH_TANGGUNGAN'])

drop_data()
dataset
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
for column in dataset:
    if dataset[column].dtypes== object:
        dataset[column]=le.fit_transform(dataset[column])
dataset
x=dataset.iloc[:, :-1].values
y=dataset.iloc[:, -1].values
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=0)
##NAIVE BAYES
from sklearn.naive_bayes import MultinomialNB
modelnb= MultinomialNB()
means = dataset.groupby(["ACTUAL_CLASS"]).mean() # Estimate me
an of each class, feature
var = dataset.groupby(["ACTUAL_CLASS"]).var() # Estimate varia
nce of each class, feature
prior = (dataset.groupby("ACTUAL_CLASS").count() / len(dataset
)).iloc[:, -1] # Estimate prior probabilities
classes = np.unique(dataset["ACTUAL_CLASS"].tolist()) # Storin
g all possible classes

prior
##DECISSION TREE
from sklearn import tree
```



```

modeltr=tree.DecisionTreeClassifier(random_state=0, max_depth=
None, min_samples_split=2, min_samples_leaf=1, min_weight_frac
tion_leaf=0, max_leaf_nodes=None, min_impurity_decrease=0)
clfnb = modelnb.fit(x_train, y_train)
clftr = modeltr.fit(x_train, y_train)

y_pred_nb= clfnb.predict(x_test)
y_pred_tr= clftr.predict(x_test)
y_pred_nb
y_pred_tr
from sklearn.metrics import confusion_matrix
cm_nb = confusion_matrix(y_test, y_pred_nb)
cm_tr = confusion_matrix(y_test, y_pred_tr)
print ('Hasil Confusion Matriks dengan Metode Naive Bayes')
cm_nb
print ('Hasil Confusion Matriks dengan Metode Decision tree')
cm_tr
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
akurasi_nb = classification_report(y_test, y_pred_nb)
akurasi_tr = classification_report(y_test, y_pred_tr)

print ('Hasil Akurasi, Presisi, Recall, dan F1-
score dengan Metode Naive Bayes')
print (akurasi_nb)

print ('Hasil Akurasi, Presisi, Recall, dan F1-
score dengan Metode Decision Tree')
print (akurasi_tr)
akurasi_nb = accuracy_score (y_test, y_pred_nb)
print ('Tingkat akurasi dengan menggunakan NB :%d persen'%(aku
rasi_nb*100))

akurasi_tr = accuracy_score (y_test, y_pred_tr)
print ('Tingkat akurasi dengan menggunakan DT :%d persen'%(aku
rasi_tr*100))
import graphviz as gv
dot_data= tree.export_graphviz(modeltr, out_file=None, feature
_names=dataset.columns[:-1], max_depth= 4, filled= True)
graph = gv.Source(dot_data)
graph.view()
graph

```

5 - 200605220015 - Prediksi Kondisi Pembayaran Biaya Pendidikan Sekolah Berbasis Klasifikasi Ekonomi Siswa

by Dini Kristianti

Submission date: 28-Dec-2022 11:39AM (UTC+0700)

Submission ID: 1987024367

File name: DINI_KRISTIANI_THESIS_FIKS_after_sidang_fiks_bismillah_oke.pdf (2.08M)

Word count: 27690

Character count: 152812

5 - 200605220015 - Prediksi Kondisi Pembayaran Biaya Pendidikan Sekolah Berbasis Klasifikasi Ekonomi Siswa

ORIGINALITY REPORT

19%

SIMILARITY INDEX

18%

INTERNET SOURCES

11%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2%
2	pdfs.semanticscholar.org Internet Source	2%
3	www.scribd.com Internet Source	1%
4	repository.dinamika.ac.id Internet Source	1%
5	repository.its.ac.id Internet Source	1%
6	lib.unnes.ac.id Internet Source	1%
7	Haditsah Annur. "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes", ILKOM Jurnal Ilmiah, 2018 Publication	1%
8	repository.ub.ac.id Internet Source	<1%

9	repository.ibik.ac.id Internet Source	<1 %
10	jurnal.akba.ac.id Internet Source	<1 %
11	repository.uinjkt.ac.id Internet Source	<1 %
12	Alman Muhammadin, Irwan Agus Sobari. "ANALISIS SENTIMEN PADA ULASAN APLIKASI KREDIVO DENGAN ALGORITMA SVM DAN NBC", Reputasi: Jurnal Rekayasa Perangkat Lunak, 2021 Publication	<1 %
13	www.gamedia.com Internet Source	<1 %
14	achmadzulfie.blogspot.co.id Internet Source	<1 %
15	ejournal.upbatam.ac.id Internet Source	<1 %
16	Submitted to Purdue University Student Paper	<1 %
17	download.garuda.kemdikbud.go.id Internet Source	<1 %
18	faridaniva.blogspot.com Internet Source	<1 %

19	id.scribd.com Internet Source	<1 %
20	Mohammad Wartaka, Anita Sitio. "SEGMENTASI PEMILIHAN PERGURUAN TINGGI SWASTA (PTS) DENGAN MENGUNAKAN TEKNIK ANALISIS KONJOIN", Jurnal Ilmiah Binaniaga, 2018 Publication	<1 %
21	Submitted to Sriwijaya University Student Paper	<1 %
22	fe-akuntansi.unila.ac.id Internet Source	<1 %
23	id.123dok.com Internet Source	<1 %
24	e-journal.janabadra.ac.id Internet Source	<1 %
25	de.scribd.com Internet Source	<1 %
26	journal.ugm.ac.id Internet Source	<1 %
27	text-id.123dok.com Internet Source	<1 %
28	jurnal.una.ac.id Internet Source	<1 %

29	library.palcomtech.com Internet Source	<1 %
30	ejournal.nusamandiri.ac.id Internet Source	<1 %
31	media.neliti.com Internet Source	<1 %
32	www.facebook.com Internet Source	<1 %
33	digilib.uin-suka.ac.id Internet Source	<1 %
34	Submitted to Universitas Diponegoro Student Paper	<1 %
35	20525940.siap-sekolah.com Internet Source	<1 %
36	Submitted to IAIN Batusangkar Student Paper	<1 %
37	123dok.com Internet Source	<1 %
38	jurnal.darmajaya.ac.id Internet Source	<1 %
39	ojs.stmik-banjarbaru.ac.id Internet Source	<1 %
40	contohaku1.blogspot.com Internet Source	<1 %

41	Frandy Ristiano, Nurmalasari Nurmalasari, Ani Yoraeni. "Impementasi Metode Naive Bayes Untuk Prediksi Harga Emas", Computer Science (CO-SCIENCE), 2021 Publication	<1 %
42	repository.uin-suska.ac.id Internet Source	<1 %
43	Submitted to Sultan Agung Islamic University Student Paper	<1 %
44	dspace.uii.ac.id Internet Source	<1 %
45	repository.teknokrat.ac.id Internet Source	<1 %
46	repository.unair.ac.id Internet Source	<1 %
47	repository.unhas.ac.id Internet Source	<1 %
48	moam.info Internet Source	<1 %
49	repository.unib.ac.id Internet Source	<1 %
50	Ratih Yulia Hayuningtyas. "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita", Jurnal Informatika, 2019 Publication	<1 %

51 Fatkhan Amirul Huda, Munawar Thoharuddin, Avelius Dominggus Sore. "PENGARUH KONDISI SOSIAL EKONOMI ORANG TUA TERHADAP KESIAPAN KERJA SISWA SMK KEAHLIAN TEKNIK KOMPUTER DAN JARINGAN SE-KOTA SINTANG", VOX EDUKASI: Jurnal Ilmiah Ilmu Pendidikan, 2019
Publication <1 %

52 Submitted to Universitas Muria Kudus
Student Paper <1 %

53 sir.stikom.edu
Internet Source <1 %

54 etheses.uin-malang.ac.id
Internet Source <1 %

55 islamhouse.com
Internet Source <1 %

56 pt.scribd.com
Internet Source <1 %

57 ojs.amikom.ac.id
Internet Source <1 %

58 digilib.uinsby.ac.id
Internet Source <1 %

59 1library.org
Internet Source <1 %

60 adoc.pub

Internet Source

<1 %

61

fokammsi.wordpress.com

Internet Source

<1 %

62

fr.scribd.com

Internet Source

<1 %

63

repository.ipb.ac.id

Internet Source

<1 %

64

trusteelink.blogspot.com

Internet Source

<1 %

65

www.coursehero.com

Internet Source

<1 %

66

www.laduni.id

Internet Source

<1 %

67

journal.institutpendidikan.ac.id

Internet Source

<1 %

68

repository.upi.edu

Internet Source

<1 %

69

Oki Rahadianto Sutopo. "Hidup adalah Perjuangan: Strategi Pemuda Yogyakarta dalam Menghadapi Transisi dari Pendidikan ke Kerja", MASYARAKAT: Jurnal Sosiologi, 2015

Publication

<1 %

70

iopscience.iop.org

Internet Source

<1 %

71	jurnalunibi.unibi.ac.id Internet Source	<1 %
72	Submitted to President University Student Paper	<1 %
73	seminar.ilkom.unsri.ac.id Internet Source	<1 %
74	www.repository.uinjkt.ac.id Internet Source	<1 %
75	eprints.undip.ac.id Internet Source	<1 %
76	Submitted to Universitas Respati Indonesia Student Paper	<1 %
77	eprints.uns.ac.id Internet Source	<1 %
78	Submitted to Tarumanagara University Student Paper	<1 %
79	Submitted to Universitas Putera Indonesia YPTK Padang Student Paper	<1 %
80	core.ac.uk Internet Source	<1 %
81	e-journal.uajy.ac.id Internet Source	<1 %
82	ejurnal.ars.ac.id Internet Source	<1 %

<1 %

83

es.scribd.com

Internet Source

<1 %

84

jurnal.pancabudi.ac.id

Internet Source

<1 %

85

repository.unikal.ac.id

Internet Source

<1 %

86

repository.unri.ac.id

Internet Source

<1 %

87

vaskoedo.wordpress.com

Internet Source

<1 %

88

digilib.iain-palangkaraya.ac.id

Internet Source

<1 %

89

fexdoc.com

Internet Source

<1 %

90

pejoeangtoga.blogspot.com

Internet Source

<1 %

91

repository.usahidsolo.ac.id

Internet Source

<1 %

92

stikomtb.ac.id

Internet Source

<1 %

93

www.brainytutorial.net

Internet Source

<1 %

94	Romindo Romindo. "Implementasi Metode SAW Terhadap Sistem Pendukung Keputusan Memilih Dosen Terbaik Pada Politeknik Ganesha", remik, 2020 Publication	<1 %
95	Yulianus Palopak, Zefanya Lumbantobing. "Analisa dan Perancangan Sistem Informasi Pendukung Keputusan Berbasis Web Untuk Pemilihan Penghuni Asrama UNAI Menggunakan Metode SAW", TelKa, 2019 Publication	<1 %
96	digilibadmin.unismuh.ac.id Internet Source	<1 %
97	ejurnal.undana.ac.id Internet Source	<1 %
98	johannessimatupang.wordpress.com Internet Source	<1 %
99	mvdw.org Internet Source	<1 %
100	repository.usd.ac.id Internet Source	<1 %
101	www.bantenprov.go.id Internet Source	<1 %
102	Rahmat Robi Waliyansyah, Citra Fitriyah. "Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-	<1 %

Nearest Neighbor (k-NN)", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2019

Publication

103	Submitted to Universitas Brawijaya Student Paper	<1 %
104	ecampus.iainbatusangkar.ac.id Internet Source	<1 %
105	ejournal-binainsani.ac.id Internet Source	<1 %
106	repository.uinsu.ac.id Internet Source	<1 %
107	repository.unej.ac.id Internet Source	<1 %
108	www.ekarir.com Internet Source	<1 %
109	Anggita Tri Prawasti, Muhammad Abduh. "Upaya Guru dalam Memotivasi Siswa Sekolah Dasar dengan Status Sosial Ekonomi Rendah", Jurnal Basicedu, 2022 Publication	<1 %
110	Haryo Bagus Setyawan. "Sistem Pendukung Keputusan Seleksi Beasiswa Bantuan Belajar Mahasiswa Menggunakan Metode Modified-Simple Additive Weighting", Jurnal Sistem dan Informatika (JSI), 2020 Publication	<1 %

111 Mohammad Farid Naufal, Selvia Ferdiana Kusuma, Kevin Christian Tanus, Raynaldy Valentino Sukiwun et al. "Analisis Perbandingan Algoritma Klasifikasi Citra Chest X-ray Untuk Deteksi Covid-19", Teknika, 2021
Publication

<1 %

112 Oman Somantri, Dairoh Dairoh. "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2019
Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off