

**KLASIFIKASI KETERLAMBATAN PEMBAYARAN SPP
MENGUNAKAN *NAÏVE BAYES* DAN
*SUPPORT VECTOR MACHINE***

THESIS

**Oleh:
HUZAIFAH UMAR
NIM. 210605210001**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PENGANTAR

**KLASIFIKASI KETERLAMBATAN PEMBAYARAN SPP
MENGUNAKAN *NAÏVE BAYES* DAN
*SUPPORT VECTOR MACHINE***

THESIS

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

**Oleh:
HUZAIFAH UMAR
NIM. 210605210001**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PENGANTAR

**KLASIFIKASI KETERLAMBATAN PEMBAYARAN SPP
MENGUNAKAN *NAÏVE BAYES* DAN
*SUPPORT VECTOR MACHINE***

THESIS

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

**Oleh:
HUZAIFAH UMAR
NIM. 210605210001**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

KLASIFIKASI KETERLAMBATAN PEMBAYARAN SPP MENGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

THESIS

Oleh:

HUZAIFAH UMAR

NIM. 210605210001

Telah diperiksa dan disetujui untuk diuji:

Tanggal: 01 Maret 2024

Pembimbing I,



Dr. Ririen Kusumawati, S.Si., M.Kom

NIP. 19720309 200501 2 002

Pembimbing II,



Dr. M. Imamudin, Lc., MA

NIP. 19740602 200901 1 010

Mengetahui,

Ketua Program Studi Magister Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Arifto Crysdiyan

NIP. 198414 200901 1 008

HALAMAN PENGESAHAN

KLASIFIKASI KETERLAMBATAN PEMBAYARAN SPP MENGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

THESIS

Oleh:
HUZAIFAH UMAR
NIM. 210605210001

Telah Dipertahankan di Depan Dewan Penguji Thesis
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 01 Maret 2024

Susunan Dewan Penguji

Penguji Utama : Dr. M. Amin Hariyadi, MT
NIP. 19670118 200501 1 001

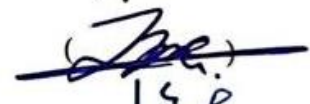
Ketua Penguji : Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

Sekretaris Penguji : Dr. Ririen Kusumawati, S.Si., M.Kom
NIP. 19720309 200501 2 002

Anggota Penguji : Dr. M. Imamudin, Lc, MA
NIP. 19740602 200901 1 010

Tanda Tangan

()


()

()

()

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




Arif Crysdiyan
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Huzaiyah Umar

NIM : 210605210001

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa thesis yang saya tulis ini benar benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 01 Maret 2024

Yang membuat pernyataan,



Huzaiyah Umar

NIM. 210605210001

MOTTO

“Tuntutlah ilmu sampai engkau mampu menginjakan kakimu
di surga Allah Subhanahu Wata’alah”

PERSEMBAHAN

Alhamdulillah puji syukur penulis panjatkan kehadiran Allah Subhanahu Wata'alah atas karunia rahmat, taufik dan hidayah-Nya sehingga penulis masih diberikan kesempatan untuk menyelesaikan thesis ini yang digunakan sebagai salah satu syarat untuk memperoleh gelar Magister Komputer (M.Kom) di Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Sholawat serta salam penulis haturkan kepada baginda Rasulullah Muhammad Shallallahu'alaihi Wasallam atas seluruh perjuangan beliau dalam menegakkan agama Islam sebagai Rahmatan Lil 'Aalamin. Thesis ini penulis persembahkan untuk:

1. Allah Subhanahu Wata'alah dan baginda Rasulullah Muhammad Shallallahu'alaihi Wasallam.
2. Kedua orang tua tercinta, ibunda Fatimah Dahlan dan ayahanda Umar Halipan yang selalu menasehati, memotivasi, mendoakan, dan mendukung agar thesis ini bisa diselesaikan dengan baik.
3. Seluruh dosen Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan ilmunya kepada saya.
4. Saudara dan saudari saya yang turut mendoakan dan memberikan dukungan.
5. Seluruh rekan mahasiswa satu angkatan Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan saran dan kritik yang membangun pada thesis ini.
6. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah berpartisipasi dalam menyelesaikan Thesis ini.

KATA PENGANTAR

Alhamdulillah segala puji hanya milik Allah Subhanahu Wata'alah atas karunia rahmat, taufik dan hidayah-Nya sehingga penulis masih diberikan kesempatan untuk menyelesaikan thesis ini. Sholawat serta salam selalu penulis haturkan kepada baginda Rasulullah Muhammad Shallallahu'alaihi Wasallam atas segala perjuangannya dalam menegakkan agama Islam sebagai Rahmatan Lil 'Aalamin.

Penulis menyadari bahwa dalam penyusunan thesis ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan ucapan terima kasih kepada:

1. Allah Subhanahu Wata'alah dan baginda Rasulullah Muhammad Shallallahu'alaihi Wasallam.
2. Kedua orang tua tercinta, ibunda Fatimah Dahlan dan ayahanda Umar Halipan yang selalu menasehati, memotivasi, mendoakan, dan mendukung agar thesis ini bisa diselesaikan dengan baik.
3. Bapak Prof. Dr. H. M. Zainuddin, MA., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang
4. Ibu Dr. Sri Harini, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
5. Bapak Dr. Cahyo Crysdiyan, selaku Ketua Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang
6. Ibu Dr. Ririen Kusumawati, S.Si., M.Kom dan Bapak Dr. M. Imamudin, Lc., MA selaku Dosen Pembimbing yang telah meluangkan waktunya dalam memberikan bimbingan dengan penuh kesabaran.
7. Bapak Dr. M. Amin Hariyadi, MT dan Bapak Dr. Zainal Abidin, M.Kom selaku Dosen Penguji yang telah memberikan saran dan kritik yang membangun pada thesis ini.
8. Seluruh dosen Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan ilmunya kepada saya.
9. Saudara dan saudari saya yang turut mendoakan dan memberikan dukungan.

10. Kak Syukur Halim dan Kak Rufaida Wati yang selalu memberikan motivasi dan dukungan kepada penulis.
11. Mas Ainur Rohman, Mas Andryan dan Mas Diky yang telah bersedia membantu penulis dalam menyelesaikan thesis ini.
12. Seluruh rekan mahasiswa satu angkatan Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan saran dan kritik yang membangun pada thesis ini.
13. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah berpartisipasi dalam menyelesaikan Thesis ini.

Malang, 01 Maret 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
PERNYATAAN KEASLIAN TULISAN	vi
MOTTO	vii
PERSEMBAHAN	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
Abstrak	xv
Abstract	xvi
مستخلص البحث	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Pernyataan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah	5
1.6 Sistematika Penulisan.....	6
BAB II STUDI PUSTAKA.....	7
2.1 Klasifikasi Keterlambatan Pembayaran.....	7
2.2 Kerangka Teori	10
BAB III METODOLOGI PENELITIAN.....	14
3.1 Tahapan Penelitian	14
3.1.1 Perencanaan	14
3.1.2 Pelaksanaan	20
3.1.3 Pengamatan	21
BAB IV NAÏVE BAYES	22
4.1 Desain.....	22
4.2 Uji Coba	30
4.2.1 Percobaan Pertama (Split Data 90:10).....	31
4.2.2 Percobaan Kedua (Split Data 80:20).....	33
4.2.3 Percobaan Ketiga (Split Data 70:30).....	34
4.2.4 Percobaan Keempat (Split Data 60:40).....	36
4.2.5 Percobaan Kelima (Split Data 50:50).....	37

4.3	Performa Naïve Bayes	38
BAB V	<i>SUPPORT VECTOR MACHINE</i>	40
5.1	Desain	40
5.2	Uji Coba	41
5.2.1	Percobaan Pertama (Split Data 90:10).....	42
5.2.2	Percobaan Kedua (Split Data 80:20)	43
5.2.3	Percobaan Ketiga (Split Data 70:30).....	45
5.2.4	Percobaan Keempat (Split Data 60:40).....	46
5.2.5	Percobaan Kelima (Split Data 50:50).....	47
5.3	Performa <i>Support Vector Machine</i>	49
BAB VI	PEMBAHASAN	50
6.1.	Hasil Pengujian NB dan SVM.....	51
6.2.	Perbandingan Performa NB dan SVM	53
6.3.	Korelasi Penelitian Dengan Nilai Nilai Islam	55
BAB VII	PENUTUP	59
7.1	Kesimpulan.....	59
7.2	Saran	59
DAFTAR PUSTAKA	60

DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori.....	11
Gambar 3.1 Tahapan Penelitian.....	14
Gambar 3.2 Desain Sistem.....	20
Gambar 3.3 Confusion Matrix	21
Gambar 4.1 Desain Alur Naive Bayes	22
Gambar 5.1 Desain Alur SVM.....	40
Gambar 6.1 Performa Naive Bayes.....	51
Gambar 6.2 Performa Support Vector Machine	52
Gambar 6.3 Perbandingan NB dan SVM.....	53
Gambar 6.4 Rata Rata Performa NB dan SVM	55

DAFTAR TABEL

Tabel 2.1 Daftar Penelitian Terdahulu	12
Tabel 3.1 Data Siswa	15
Tabel 3.2 Data Orang Tua.....	15
Tabel 3.3 Kategori Pembayaran SPP	16
Tabel 3.4 Integrasi Data	17
Tabel 3.5 Kategori Pekerjaan Orang Tua.....	18
Tabel 3.6 Kategori Penghasilan Orang Tua	19
Tabel 3.7 Kategori Jumlah Tanggungan	19
Tabel 3.8 Dataset	20
Tabel 4.1 Data Training	23
Tabel 4.2 Probabilitas Pekerjaan Ayah	24
Tabel 4.3 Probabilitas Pekerjaan Ibu	25
Tabel 4.4 Probabilitas Penghasilan Ayah	27
Tabel 4.5 Probabilitas Penghasilan Ibu.....	28
Tabel 4.6 Probabilitas Jumlah Tanggungan	28
Tabel 4.7 Data Testing	29
Tabel 4.8 Hasil Klasifikasi.....	30
Tabel 4.9 Uji Coba Percobaan Pertama	31
Tabel 4.10 Uji Coba Percobaan Kedua	33
Tabel 4.11 Uji Coba Percobaan Ketiga.....	34
Tabel 4.12 Uji Coba Percobaan Keempat.....	36
Tabel 4.13 Uji Coba Percobaan Kelima.....	37
Tabel 4.14 Performa Naive Bayes	39
Tabel 5.1 Uji Coba Percobaan Pertama	42
Tabel 5.2 Uji Coba Percobaan Pertama	43
Tabel 5.3 Uji Coba Percobaan Pertama	45
Tabel 5.4 Uji Coba Percobaan Pertama	46
Tabel 5.5 Uji Coba Percobaan Pertama	48
Tabel 5.6 Performa Support Vector Machine	49

Abstrak

Umar, Huzaifah. 2024. **Klasifikasi Keterlambatan Pembayaran SPP Menggunakan *Naive Bayes* dan *Support Vector Machine***. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Ririen Kusumawati, S.Si., M.Kom, (II) Dr. M. Imamudin, Lc, MA

Kata Kunci: Klasifikasi, SPP, *Naive Bayes*, *Support Vector Machine*

Keterlambatan pembayaran SPP merupakan salah satu masalah yang sering dihadapi oleh sekolah. Hal ini dikarenakan dapat berpengaruh pada proses penggajian guru dan karyawan maupun program peningkatan mutu pendidikan di lingkungan sekolah. Penelitian ini bertujuan untuk melakukan klasifikasi keterlambatan pembayaran SPP menggunakan *Naive Bayes* dan *Support Vector Machine*. Uji coba metode dilakukan dengan 5 kali percobaan. Berdasarkan hasil pengujian diperoleh rata rata performa *Naive Bayes* yaitu *accuracy* (62,88%), *precision* (65,27%), *recall* (77,42%) dan *f1-score* (70,75%). Sedangkan rata rata performa *Support Vector Machine* yaitu *accuracy* (63,51%), *precision* (62,25%), *recall* (94,48%) dan *f1-score* (75,04%).

Abstract

Umar, Huzaifah. 2024. *The Clasification of Late School Fee Payment Using Naive Bayes and Support Vector Machine*. Thesis. Magister of Informatics. Faculty of Science and Technology. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Dr. Ririen Kusumawati, S.Si., M.Kom, (II) Dr. M. Imamudin, Lc, MA

Keywords: Classification, School Fee, Naïve Bayes, Support Vector Machine

Late fee payment often occurs at schools. It can affect teacher and staff salary payment process and educational quality improvement programs at school. The research aims to classify the late school fee payment using Naïve Bayes and Support Vector Machine. The researcher conducted the method trial five times. The testing result shows that the Naïve Bayes average performances for accuracy, precision, recall, and f1-score are 62.88%, 65.27%, 77.42%, and 70.75%, respectively. Meanwhile, the average performances of the Support Vector Machine for accuracy, precision, recall, and f1-score are 63.51%, 62.25%, 94.48%, and 75.04%.

مستخلص البحث

عمر، حذيفة. 2024. تصنيف التأخر في سداد الرسوم الدراسية باستخدام ساذج بايز وآلة المتجه الداعم. رسالة الماجستير. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. ريرين كوسوماواتي، الماجستير. المشرف الثاني: د. محمد إمام الدين، الماجستير.

الكلمات الرئيسية: تصنيف، رسوم دراسية، ساذج بايز، آلة المتجه الداعم.

يعد التأخر في سداد الرسوم الدراسية أحد المشكلات التي تواجهها المدارس في كثير من الأحيان. هذا لأنه يمكن أن يؤثر على عملية كشف المرتبات للمعلمين والموظفين وكذلك برامج تحسين جودة التعليم في البيئة المدرسية. تهدف هذه الرسالة إلى تصنيف التأخر في سداد الرسوم الدراسية باستخدام ساذج بايز وآلة المتجه الداعم. أجريت تجربة الطريقة بخمس تجارب. بناء على نتائج الاختبار، كان متوسط أداء ساذج بايز هو الدقة (62.88%) والثبات (65.27%) والاستدعاء (77.42%) ودرجة ف1 (70.75%). في حين أن متوسط أداء آلة المتجه الداعم هو الدقة (63.51%) والثبات (62.25%) والاستدعاء (94.48%) ودرجة ف1 (75.04%).

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan memiliki peran sebagai fondasi pembangunan kepribadian manusia. Dengan adanya pendidikan seseorang akan mampu mengatasi masalah kebodohan yang melanda pribadinya bahkan dalam lingkup yang lebih luas yaitu mencerdaskan kehidupan bangsa. Sebagaimana amanat Undang Undang Dasar (UUD 1945) Negara Republik Indonesia, mencerdaskan kehidupan bangsa merupakan tujuan yang hendak diwujudkan oleh negara. Penyelenggaraan pendidikan merupakan bagian dari cara untuk mencerdaskan kehidupan bangsa.

Dalam konteks penyelenggaraan pendidikan, salah satu bagian terpenting adalah tersedianya pembiayaan pendidikan. Dalam penelitian Noor Islahudin & Ramadhani Wulandari (2022) pembiayaan pendidikan memiliki hubungan erat dengan mutu pendidikan yang baik. Menurut Arwildayanto *et al.*, (2017) faktor yang memiliki keterkaitan secara langsung dalam menjalankan proses pendidikan adalah pembiayaan dan mutu pendidikan.

MTs Lukmanul Hakim merupakan lembaga pendidikan swasta di Samarinda yang pembiayaan pendidikannya dibebankan kepada siswa. Salah satu biaya pendidikan yang harus dipenuhi adalah sumbangan pembinaan pendidikan atau yang biasa dikenal dengan SPP. Menurut Bahtiar & Firmansyah (2017) pada umumnya SPP diterapkan oleh sekolah swasta karena dalam pengelolaan pendidikannya dibebankan kepada masyarakat sedangkan sekolah negeri biaya pengelolaan pendidikannya dibiayai oleh pemerintah.

Masalah yang terjadi pada pembayaran SPP adalah apabila siswa terlambat membayar tagihan SPP dari tenggat waktu yang telah diberikan oleh pihak sekolah. Hal ini akan bermasalah karena keterlambatan pembayaran SPP dapat berpengaruh pada proses penggajian guru dan karyawan maupun program peningkatan mutu pendidikan di lingkungan sekolah. Berdasarkan data yang bersumber dari bendahara sekolah diketahui bahwa pada tahun ajaran 2023/2024 jumlah siswa yang terlambat membayar SPP sebesar 42% dari total keseluruhan siswa yang berjumlah 248 siswa. Oleh karena itu, perlu adanya solusi terhadap permasalahan tersebut sehingga dapat menjadi bahan evaluasi oleh pihak sekolah. Solusi yang dapat dilakukan adalah dengan melakukan klasifikasi keterlambatan pembayaran SPP.

Menurut Wibowo (2017) klasifikasi merupakan suatu pengelompokan data dimana data yang digunakan memiliki target atau kelas label. Konsep pengelompokan data atau klasifikasi dapat ditemukan dalam surah Al An'am ayat 142 yaitu: **وَمِنَ الْأَنْعَامِ حَمُولَةٌ وَفَرَشَاءٌ كُلُوا مِمَّا رَزَقَكُمُ اللَّهُ وَلَا تَتَّبِعُوا خُطُوَاتِ الشَّيْطَانِ إِنَّهُ لَكُمْ عَدُوٌّ مُّبِينٌ** artinya: *“dan di antara hewan-hewan ternak itu ada yang dijadikan pengangkut beban dan ada (pula) yang untuk disembelih. Makanlah rezeki yang diberikan Allah kepadamu dan janganlah kamu mengikuti langkah-langkah setan. Sesungguhnya setan itu musuh yang nyata bagimu.”*

Menurut tafsir Ibnu Katsir yang diterjemahkan oleh Ghoffar E.M. *et al.*, (2004) diceritakan bahwa Allah menjadikan binatang ternak yang sebagian darinya dapat dijadikan sebagai kendaraan angkutan, dan sebagian yang lain dijadikan sebagai hewan potong untuk manusia. Sebagaimana yang dikatakan oleh As-Sauri, dari Abu Ishaq dari Abul Ahwas, dari Abdullah berkaitan dengan makna firman Allah: maksud dari *Hamuulah* adalah binatang unta yang dijadikan sebagai kendaraan

angkutan, sedangkan *Farsy* adalah unta yang masih muda. Demikian menurut riwayat Imam Hakim yang mengatakan sanad asar ini sahih, tetapi Imam Bukhari dan Muslim tidak menyetujuinya.

Ibnu Abbas mengatakan bahwa *Hamuulah* adalah unta yang sudah dewasa, sedangkan *Farsy* adalah unta yang masih muda. Adapun yang dimaksud “*makanlah dari rezeki yang telah diberikan Allah kepada kalian*” yaitu rezeki berupa buah-buahan, hasil tanaman, dan binatang ternak. Sedangkan kalimat “*Dan janganlah kalian mengikuti langkah-langkah setan, sesungguhnya setan itu musuh yang nyata bagimu*” yaitu jangan mengikuti jalan yang dianjurkan oleh setan, sebagaimana yang ditempuh oleh orang-orang musyrik yang berani mengharamkan buah-buahan dan hasil tanaman yang direzekikan oleh Allah buat mereka sebagai buah-buahan mereka yang mereka nisbatkan kepada Allah.

Beberapa penelitian tentang klasifikasi pembayaran SPP telah dilakukan seperti penelitian Bahtiar & Firmansyah (2017) tentang klasifikasi ketepatan waktu pembayaran SPP menggunakan metode *Naïve Bayes* pada ponpes Al Arifah. Dari hasil penelitian memperoleh nilai akurasi sebesar 63,64%.

Penelitian sejenis juga pernah dilakukan oleh Prabowo *et al.*, (2021). Keterlambatan pembayaran SPP menjadi masalah yang serius karena SPP menjadi sumber dana yang digunakan untuk mendukung penyelenggaraan pendidikan di lingkungan pesantren. Penelitian ini juga menggunakan algoritma *Naïve Bayes*. Berdasarkan hasil perhitungan diperoleh nilai akurasi sebesar 63,64%.

Kemudian penelitian yang dilakukan Muqorobin *et al.*, (2019) mengatakan bahwa pada tahun ajaran 2017/2018 siswa terlambat membayar SPP berjumlah

60% dari total keseluruhan siswa yang berjumlah 236. Penelitian ini menggabungkan algoritma *Naïve Bayes* dengan *Information Gain* dan. Hasil yang diperoleh sebesar 90% sedangkan jika hanya menggunakan *Naive Bayes* hanya memperoleh nilai akurasi sebesar 80%.

Penelitian Rosiana (2022) menggunakan model *Support Vector Machine* dalam menganalisis kemungkinan keterlambatan pembayaran SPP pada SMP Perintis 2 Bandar Lampung. Hasil penelitian menyatakan *Support Vector Machine* dapat digunakan sebagai salah satu model dalam melakukan klasifikasi keterlambatan pembayaran.

Penelitian Nugraha (2021) yang melakukan klasifikasi penyesuaian UKT mahasiswa terdampak pandemi Covid-19 menggunakan algoritma *Support Vector Machine*. Dari hasil pengujian diperoleh nilai akurasi sebesar 83,00% dari 100 dataset yang diujicobakan.

Penelitian Widayani & Harliana (2021) melakukan klasifikasi penundaan biaya kuliah mahasiswa dengan algoritma *Support Vector Machine*. Penelitian ini bertujuan untuk mendapatkan nilai akurasi tertinggi pada algoritma *Support Vector Machine* dalam memberikan rekomendasi penundaan pembayaran kuliah. Dari hasil pengujian diperoleh tingkat akurasi tertinggi didapat pada skenario keempat yaitu sebesar 87%.

Berdasarkan berbagai penelitian yang telah diuraikan dapat disimpulkan bahwa algoritma *Naïve Bayes* dan *Support Vector Machine* dapat digunakan dalam melakukan klasifikasi keterlambatan pembayaran SPP. Dalam penelitian Wu *et al.*, (2008) kedua metode ini masuk dalam daftar 10 algoritma terbaik dalam *data*

mining berdasarkan paper yang dipublikasikan pada tahun 2006 oleh *IEEE International Conference on Data mining*.

Berdasarkan uraian tersebut penelitian kali ini akan dilakukan klasifikasi keterlambatan pembayaran SPP menggunakan *Naïve Bayes* dan *Support Vector Machine*. Hasil penelitian ini diharapkan dapat dijadikan sebagai bahan pertimbangan oleh pihak sekolah dalam mengatasi keterlambatan pembayaran SPP.

1.2 Pernyataan Masalah

Pernyataan masalah pada penelitian ini adalah berapa tingkat akurasi dari *Naïve Bayes* dan *Support Vector Machine* dalam melakukan klasifikasi keterlambatan pembayaran SPP?

1.3 Tujuan Penelitian

Tujuan pada penelitian ini adalah mengetahui tingkat akurasi dari *Naïve Bayes* dan *Support Vector Machine* dalam melakukan klasifikasi keterlambatan pembayaran SPP.

1.4 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Dapat digunakan oleh pihak sekolah sebagai sumber evaluasi kinerja sekolah terutama pada bidang keuangan.
2. Dapat digunakan sebagai salah satu acuan atau pertimbangan pemangku kebijakan dalam menentukan kebijakan pembayaran SPP di sekolah.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah:

1. Data yang digunakan adalah data siswa di Madrasah Tsanawiyah Lukmanul Hakim Samarinda.
2. Data siswa yang digunakan diambil mulai dari tahun ajaran 2021/2022 sampai 2023/2024.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, pernyataan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II STUDI PUSTAKA

Bab ini memuat teori teori dan karya ilmiah yang berkaitan dengan penelitian ini yang bersumber dari beberapa referensi yang valid.

BAB III METODOLOGI PENELITIAN

Bab ini memuat tentang tahapan yang digunakan dalam penelitian ini mulai dari pengumpulan data sampai evaluasi.

BAB IV NAÏVE BAYES

Bab ini membahas tentang implementasi algoritma *Naïve Bayes*.

BAB V SUPPORT VECTOR MACHINE

Bab ini membahas tentang implementasi algoritma *Support Vector Machine*.

BAB VI PEMBAHASAN

Bab ini memuat tentang pembahasan hasil penelitian.

BAB VII PENUTUP

Bab ini merupakan bab terakhir yang berisi kesimpulan dan saran dari penulis.

BAB II

STUDI PUSTAKA

2.1 Klasifikasi Keterlambatan Pembayaran

Menurut Prabowo *et al.*, (2021) biaya pendidikan sangat berperang penting dalam penyelenggaraan pendidikan di pondok pesantren. Keterlambatan dalam melakukan pembayaran SPP akan menjadi masalah sebab SPP menjadi sumber dana yang akan digunakan pesantren dalam mendukung penyelenggaraan pendidikan. Sehingga dibutuhkan solusi dalam mengatasi masalah tersebut yaitu melakukan klasifikasi pembayaran SPP berdasarkan tingkat ketepatan siswa dalam melakukan pembayaran menggunakan algoritma *Naïve Bayes* sehingga dapat menjadi bahan evaluasi pihak pesantren. Berdasarkan hasil perhitungan yang dilakukan menghasilkan tingkat akurasi sebesar 63,64%.

Kemudian penelitian yang dilakukan oleh Bahtiar & Firmansyah (2017) mengatakan masalah yang sering terjadi pada pembayaran SPP adalah ketika para siswa terlambat dalam melakukan pembayaran dari tenggat waktu yang telah ditentukan. Hal tersebut menjadi masalah besar karena SPP menjadi faktor yang penting dalam meningkatkan kualitas pelayanan di lingkungan pesantren. Faktor faktor yang menjadi penyebab diantaranya adalah penghasilan orang tua dan kebiasaan dari santri itu sendiri. Maka diperlukan solusi terhadap permasalahan tersebut dengan melakukan pengklasifikasian pembayaran SPP. Tujuan penelitian ini ialah melakukan klasifikasi pembayaran SPP berdasarkan ketepatan waktu pembayaran dengan menggunakan algoritma *Naïve Bayes*. Nilai akurasi yang diperoleh sebesar 63,64% berdasarkan 45 data latih dan 22 data uji.

Penelitian yang dilakukan oleh Muqorobin *et al.*, (2019) mengatakan bahwa pada tahun ajaran 2017/2018 terdapat 60% dari total keseluruhan seluruh siswa yang berjumlah 236 siswa terlambat membayar SPP. Hal ini menunjukkan bahwa permasalahan keterlambatan pembayaran SPP di lingkungan sekolah masih menjadi permasalahan yang serius. Oleh karena itu diperlukan sebuah solusi berupa sistem yang dapat memperkirakan siswa yang tepat atau terlambat dalam melakukan pembayaran SPP. Penelitian ini menggabungkan algoritma *Naive Bayes* dengan *Information Gain*. Hasil yang diperoleh sebesar 90% sedangkan jika hanya menggunakan *Naive Bayes* hanya memperoleh nilai akurasi sebesar 80%.

Selanjutnya Ginting *et al.*, (2020) melakukan penelitian dengan menggunakan dataset yang sama dari penelitian Muqorobin *et al.*, (2019) kemudian mengimplementasikannya ke dalam bahasa pemrograman python. Algoritma yang digunakan adalah algoritma *Naive Bayes*. Dari hasil pengujian diperoleh tingkat akurasi *Naive Bayes* sebesar 73%.

Kemudian penelitian yang dilakukan oleh Norhalimi & Siswa (2022) melakukan optimasi seleksi fitur *Information Gain* pada algoritma *Naive Bayes* dan *K-Nearest Neighbor* untuk mengatasi permasalahan keterlambatan pembayaran biaya kuliah di Universitas Muhammadiyah Kalimantan Timur. Berdasarkan data yang diperoleh terjadi kenaikan yang sangat drastis jumlah mahasiswa yang terlambat dalam melakukan pembayaran biaya kuliah. Berdasarkan hasil perhitungan seleksi fitur *Information Gain* diperoleh 4 atribut yang sangat berpengaruh yaitu fakultas, prodi, angkatan, dan gender. Hasil evaluasi *Confusion Matrix* diperoleh tingkat akurasi sebesar 55,19% menggunakan algoritma *Naive*

Bayes dengan *Information Gain*. Sedangkan pada algoritma *K-Nearest Neighbor* dengan *Information Gain* hanya diperoleh tingkat akurasi sebesar 50,76%.

Penelitian Rosiana (2022) menggunakan algoritma *Support Vector Machine* dalam menganalisa kemungkinan keterlambatan pembayaran SPP pada SMP Perintis 2 Bandar Lampung. Berdasarkan hasil pengujian menggunakan Weka dan *Ms. Excel* mendapatkan tingkat akurasi yang sama besar namun terdapat kekurangan pada *Ms. Excel* yaitu waktu yang tidak efisien dan memiliki kemungkinan *human error*. Solusi yang ditawarkan ialah pihak sekolah dapat memberikan himbauan kepada orang tua untuk menyiapkan biaya pembayaran SPP lebih awal sebelum batas waktu pembayaran. Selain itu orang tua bisa membayar SPP dengan mekanisme cicilan.

Penelitian Abdullah *et al.*, (2019) menggunakan metode *K-Nearest Neighbor* dalam melakukan analisis keterlambatan pembayaran SPP di SMK Al-Islam Surakarta. Penelitian ini diharapkan dapat dijadikan solusi oleh pihak sekolah untuk mengatasi keterlambatan pembayaran SPP di sekolah. Nilai akurasi yang dihasilkan menggunakan algoritma *K-Nearest Neighbor* adalah 86%. Adapun saran untuk penelitian selanjutnya adalah dapat menggunakan metode klasifikasi lainnya.

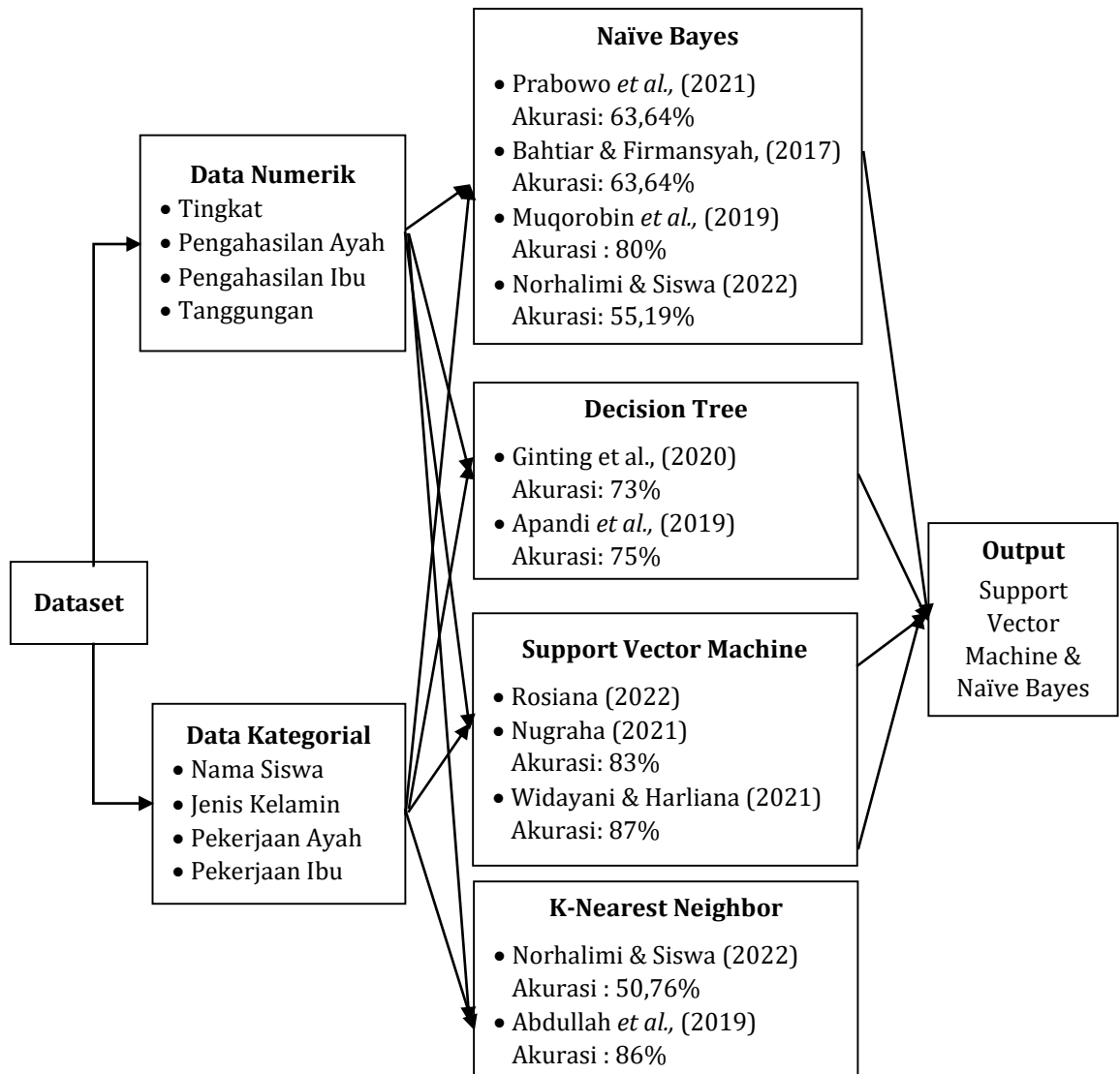
Penelitian Apandi *et al.*, (2019) menganalisa tentang kemungkinan keterlambatan pembayaran SPP menggunakan algoritma C4.5. Hasil klasifikasi dari algoritma tersebut kemudian dievaluasi dan divalidasi menggunakan *Cross Validation* dan *Confusion Matrix* agar dapat diketahui tingkat akurasinya. Berdasarkan hasil penelitian menunjukkan bahwa hasil akurasi terbaik dihasilkan oleh partisi 2 yaitu sebesar 75%.

Penelitian Nugraha (2021) melakukan klasifikasi penyesuaian UKT mahasiswa terdampak pandemi Covid-19 menggunakan algoritma *Support Vector Machine*. Dari hasil pengujian diperoleh nilai akurasi sebesar 83,00% dari 100 dataset yang diujicobakan.

Penelitian Widayani & Harliana (2021) melakukan klasifikasi penundaan biaya kuliah mahasiswa menggunakan *Support Vector Machine*. Penelitian ini bertujuan untuk mendapatkan tingkat akurasi tertinggi menggunakan algoritma *Support Vector Machine* dalam memberikan rekomendasi penundaan pembayaran kuliah. Dari hasil pengujian diperoleh tingkat akurasi tertinggi didapat pada scenario keempat yaitu sebesar 87%. Sedangkan untuk pada jenis kernel, akurasi tertinggi menggunakan kernel *Gaussian RBF* yaitu sebesar 97%.

2.2 Kerangka Teori

Kerangka teori pada penelitian ini merujuk pada penelitian relevan atau penelitian yang telah dilakukan sebelumnya sebagai bahan referensi. Kerangka teori pada penelitian ini dapat dilihat pada gambar 2.1



Gambar 2.1 Kerangka Teori

Kerangka teori pada gambar 2.1 menjelaskan tentang beberapa studi pustaka yang dilakukan peneliti dalam menyusun penelitian tentang klasifikasi keterlambatan pembayaran SPP sehingga menjadi dasar pelaksanaan penelitian ini.

- a) Data yang digunakan diperoleh dari bagian tata usaha dan bendahara MTs Lukmanul Hakim Samarinda tahun ajaran 2021/2022 sampai 2023/2024.
- b) Penelitian tentang klasifikasi keterlambatan pembayaran SPP pernah diteliti sebelumnya. Adapun daftar penelitian terdahulu dapat dilihat pada tabel 2.1

Tabel 2.1 Daftar Penelitian Terdahulu

No	Nama (Tahun)	Judul	Metode	Hasil
1	<ul style="list-style-type: none"> • Candi P. • Lukman Nul H. • Cep Lukman R. • Ade R. R. (2021) 	Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran	<i>Naïve Bayes</i>	Hasil akurasi 63,64%. Faktor yang mempengaruhi pekerjaan dan pendapatan orang tua, dan faktor lainnya
2	<ul style="list-style-type: none"> • Agus B. • Mulyawan • Suryani • Dindin F. (2017) 	Klasifikasi Ketepatan Waktu Pembayaran SPP di Pondok Pesantren Al-Arifah Menggunakan Algoritma <i>Naïve Bayes</i>	<i>Naïve Bayes</i>	Hasil akurasi 63,64%. faktor yang mempengaruhi terjadinya keterlambatan adalah ekonomi, pekerjaan dan jenis kelamin
3	<ul style="list-style-type: none"> • Muqorobin • Kusri • Emha Taufiq L. (2019) 	Optimasi Metode <i>Naïve Bayes</i> dengan Feature Selection <i>Information Gain</i> Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah	<i>Naïve Bayes + Information Gain</i>	Hasil akurasi metode <i>Naïve Bayes</i> dengan fitur <i>Information Gain</i> sebesar 90% sedangkan tidak menggunakan <i>Information Gain</i> sebesar 80%
4	<ul style="list-style-type: none"> • Victor S. G. • Kusri • Emha Taufiq L. (2020) 	Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python	Algoritma C4.5	Berdasarkan hasil penelitian diperoleh tingkat akurasi algoritma C4.5 sebesar 73%
5	<ul style="list-style-type: none"> • M. Norhalimi • Taghfirul A. S. (2022) 	Optimasi Seleksi Fitur <i>Information Gain</i> pada Algoritma <i>Naïve Bayes</i> dan <i>K-Nearest Neighbor</i>	<i>Naïve Bayes & K-NN</i>	Hasil akurasi <i>Naïve Bayes</i> dengan fitur <i>Information Gain</i> 55,19%. Sedangkan <i>K-NN</i> dengan fitur <i>Information Gain</i> 50,76%
6	<ul style="list-style-type: none"> • Dinda N. B • Agus P. W • Eka Irawan (2022) 	Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi <i>Decision Tree</i> dan <i>SVM</i>	<i>Decision Tree & SVM</i>	Diketahui faktor yang paling berpengaruh dalam keterlambatan pembayaran listrik adalah atribut tanggal penerimaan gaji
7	<ul style="list-style-type: none"> • Lela Rosiana (2022) 	Analisis Kemungkinan Keterlambatan Pembayaran SPP Menggunakan Algoritma <i>Support Vector Machine</i> (Studi Kasus: SMP Perintis 2 Bandar Lampung)	<i>SVM</i>	Tingkat akurasi dengan Weka dan <i>Ms. Excel</i> sama besar namun ada kekurangan pada <i>Ms. Excel</i> yaitu waktu tidak efisien dan kemungkinan <i>human error</i> .

Lanjutan tabel 2.1 daftar penelitian terdahulu

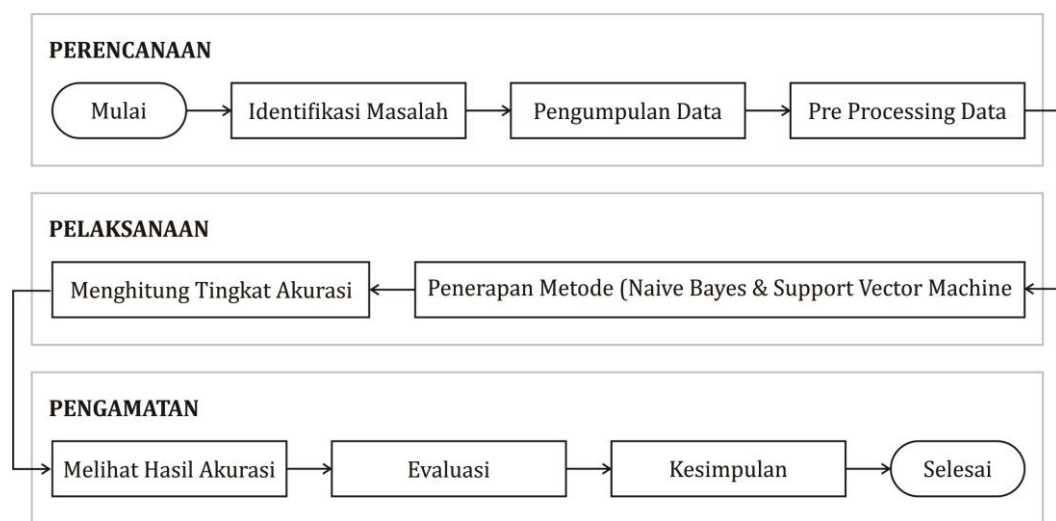
No	Nama (Tahun)	Judul	Metode	Hasil
8	<ul style="list-style-type: none"> • Tri H. Apandi • Roby B. Maulana • Rian Piarna • Dwi V. (2019) 	Menganalisis Kemungkinan Keterlambatan Pembayaran SPP dengan Algoritma C4.5 (Studi Kasus Politeknik TEDC Bandung)	Algoritma C4.5	Hasil pengujian menunjukkan nilai akurasi terbaik dihasilkan pada partisi 2 yaitu sebesar 75%
9	<ul style="list-style-type: none"> • Nugraha (2021) 	Implementasi algoritma SVM untuk klasifikasi penyesuaian uang kuliah tunggal terdampak pandemi Covid-19	SVM	Diperoleh tingkat akurasi sebesar 83,00%
10	<ul style="list-style-type: none"> • Widayani & Harliana (2021) 	Perbandingan Kernel SVM dalam melakukan klasifikasi penundaan biaya kuliah mahasiswa	SVM	Akurasi tertinggi pada skenario keempat yaitu 87%. Dan pada kernel Gaussian RBF 97%.

Berdasarkan hasil uraian dari rujukan penelitian terdahulu pada tabel 2.1 maka pada penelitian kali ini akan dilakukan klasifikasi keterlambatan pembayaran SPP menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*.

BAB III METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Berikut ini adalah tahapan penelitian yang akan dilakukan pada penelitian kali ini. Tahapan penelitian terdiri dari perencanaan, pelaksanaan dan pengamatan. Adapun desain tahapan penelitian ini dapat dilihat pada gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.1.1 Perencanaan

Langkah pertama yang dilakukan pada tahap perencanaan adalah mengidentifikasi masalah dengan menemukan latar belakang masalah, rumusan masalah dan solusi permasalahan. Selanjutnya melakukan pengumpulan data, data yang digunakan pada penelitian ini data siswa MTs Lukmanul Hakim Samarinda. Data tersebut diperoleh dari bagian tata usaha dan bendahara sekolah yang berjumlah 248 data. Data diambil mulai dari tahun ajaran 2021/2022 sampai 2023/2024. Hasil pengumpulan data dapat dilihat pada tabel 3.1, tabel 3.2, dan tabel 3.3

Tabel 3.1 Data Siswa

No	Nama	NISN	Jenis Kelamin	Tempat Lahir	Tanggal Lahir	Tingkat	Tahun Masuk
1	Siswa 1		Laki Laki	Makassar	08/01/2011	7	2023
2	Siswa 2		Laki Laki	Samarinda	14/10/2010	7	2023
3	Siswa 3		Laki Laki	Samarinda	24/09/2011	7	2023
4	Siswa 4		Laki Laki	Samarinda	20/01/2011	7	2023
5	Siswa 5		Laki Laki	Samarinda	18/06/2010	7	2023
...
244	Siswa 244		Perempuan	Balikpapan	30/11/2009	9	2021
245	Siswa 245		Laki Laki	Sulawesi USA	18/12/2007	9	2021
246	Siswa 246		Perempuan	Samarinda	26/05/2009	9	2021
247	Siswa 247		Perempuan	Balikpapan	30/12/2008	9	2021
248	Siswa 248		Laki Laki	Samarinda	13/01/2007	9	2021

Sumber: Tata Usaha Sekolah

Tabel 3.1 merupakan data siswa yang memiliki atribut seperti nama, NISN, jenis kelamin siswa, tempat dan tanggal lahir, tingkat atau kelas dan tahun masuk siswa. Atribut nama merupakan pemberian kedua orang tua sebagai identitas siswa. Atribut NISN merupakan kode pengenal siswa yang bersifat unik yang dikelola secara nasional oleh pusat daya dan statistik pendidikan dan kebudayaan kemendikbud. Atribut jenis kelamin adalah data yang mengkategorikan siswa sesuai dengan kelamin perempuan atau laki-laki. Kemudian atribut tempat dan tanggal lahir merupakan tempat dan tanggal siswa dilahirkan yang diperoleh dari akta kelahiran. Atribut tingkat atau kelas merupakan kelas siswa berada. Sedangkan atribut tahun masuk adalah tahun pertama kali siswa mendaftar di sekolah.

Tabel 3.2 Data Orang Tua

No	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah (Rp)	Penghasilan Ibu (Rp)	Tanggungan	Alamat
1	Siswa 1	Wiraswasta	Karyawan Swasta	3.000.000 - 4.999.999	3.000.000 - 4.999.999	4	Jl. Dewantara Melak Ulu
2	Siswa 2	Wiraswasta	Tidak Bekerja	3.000.000 - 4.999.999	0	2	Jl. Teuku Umar
3	Siswa 3	Wiraswasta	Ibu Rumah Tangga	1.000.000 - 2.999.999	0	3	Jl. Rajawali Dalam 4

Sumber: Tata Usaha Sekolah

Lanjutan tabel 3.2 data orang tua

No	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah (Rp)	Penghasilan Ibu (Rp)	Tanggungan	Alamat
4	Siswa 4	Tidak Bekerja	Karyawan Swasta	0	3.000.000 - 4.999.999	2	Jl Gerilya
5	Siswa 5	Karyawan Swasta	Ibu Rumah Tangg	3.000.000 – 4.999.999	0	2	Jln. AMD 59
...
246	Siswa 246	Karyawan Swasta	Tidak Bekerja	3.000.000 – 4.999.999	0	2	Jl. Wahid Hasyim
247	Siswa 247	Wirausaha	PNS/TNI/POLRI	3.000.000 – 4.999.999	3.000.000 – 4.999.999	3	Jl. Soekarno Hatta
248	Siswa 248	Nelayan	Tidak Bekerja	< 500.000	0	3	Jl. Perjuangan

Sumber: Tata Usaha Sekolah

Tabel 3.2 merupakan data orang tua siswa yang diperoleh dari tata usaha sekolah. Data tersebut memiliki atribut seperti nama, pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, tanggungan, dan alamat. Atribut nama merupakan nama siswa. Atribut pekerjaan ayah dan ibu merupakan jenis pekerjaan dari masing masing orang tua siswa. Atribut penghasilan ayah dan ibu merupakan penghasilan bulanan dari masing masing orang tua siswa. Atribut tanggungan merupakan jumlah anak yang masih menjadi tanggungan orang tua siswa.

Tabel 3.3 Kategori Pembayaran SPP

No	Nama	NISN	Jenis Kelamin	Tingkat	Kategori
1	Siswa 1		Laki Laki	7	Tepat
2	Siswa 2		Laki Laki	7	Tepat
3	Siswa 3		Laki Laki	7	Tepat
4	Siswa 4		Laki Laki	7	Tepat
5	Siswa 5		Laki Laki	7	Terlambat
...
245	Siswa 245		Laki Laki	9	Tepat
246	Siswa 246		Perempuan	9	Tepat
247	Siswa 247		Perempuan	9	Terlambat
248	Siswa 248		Laki Laki	9	Tepat

Sumber: Bendahara Sekolah

Tabel 3.3 merupakan data kategori tepat atau terlambat siswa dalam melakukan pembayaran SPP yang diperoleh dari bendahara sekolah. Berdasarkan data tersebut diketahui siswa yang dikategorikan sebagai terlambat membayar SPP sebanyak 104 siswa (42%) dan siswa yang dikategorikan sebagai tepat membayar SPP 144 siswa (58%).

Setelah data dikumpulkan, tahap selanjutnya adalah melakukan *pre processing* data. Tahap ini merupakan persiapan data awal sehingga siap digunakan pada saat permodelan. Secara garis besar tahap ini dilakukan untuk memperbaiki masalah dalam data. Kegiatan yang dilakukan adalah seleksi data, integrasi data, dan transformasi data.

1. Seleksi Data

Pada tahap ini dilakukan pemilihan atribut atau parameter yang akan digunakan dari data siswa, data orang tua dan data pembayaran SPP. Atribut yang akan digunakan adalah nama siswa, pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, jumlah tanggungan dan kategori pembayaran SPP.

2. Integrasi Data

Tahap integrasi data dilakukan dengan menggabungkan data siswa, data orang tua, dan data keterangan pembayaran sehingga menjadi satu kesatuan dataset.

Tabel 3.4 Integrasi Data

No	Nama	Pekerjaa n Ayah	Pekerjaa n Ibu	Penghasila n Ayah (Rp)	Penghasilan Ibu (Rp)	Tangg ungan	Kategori
1	Siswa 1	Wiraswas ta	Karyawan Swasta	3.000.000 - 4.999.999	3.000.000 - 4.999.999	4	Tepat
2	Siswa 2	Wiraswas ta	Tidak Bekerja	3.000.000 - 4.999.999	0	2	Tepat
3	Siswa 3	Wiraswas ta	Ibu Rumah Tangg	1.000.000 - 2.999.999	0	3	Tepat

Sumber: Tata Usaha dan Bendahara Sekolah

Lanjutan tabel 3.4 integrasi data

No	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah (Rp)	Penghasilan Ibu (Rp)	Tanggungan	Kategori
4	Siswa 4	Tidak Bekerja	Karyawan Swasta	0	3.000.000 - 4.999.999	2	Tepat
5	Siswa 5	Karyawan Swasta	Ibu Rumah Tangga	3.000.000 – 4.999.999	0	2	Terlambat
...
246	Siswa 246	Karyawan Swasta	Tidak Bekerja	3.000.000 – 4.999.999	0	2	Tepat
247	Siswa 247	Wirasaha	PNS/TNI/POLRI	3.000.000 – 4.999.999	3.000.000 – 4.999.999	3	Terlambat
248	Siswa 248	Nelayan	Tidak Bekerja	< 500.000	0	3	Tepat

Sumber: Tata Usaha dan Bendahara Sekolah

3. Transformasi Data

Pada tahap ini dilakukan konversi atribut yang telah dipilih agar dapat diproses dengan baik pada tahap modeling.

Tabel 3.5 Kategori Pekerjaan Orang Tua

Kriteria Pekerjaan	Parameter
Wiraswasta	W
Karyawan Swasta	KS
Guru/Dosen	GD
PNS/TNI/Polri	PTP
Buruh	B
Petani	P
Nelayan	N
Supir	S
Bidan Desa	BD
Pedagang	PG
Ibu Rumah Tangga	IRT
Tidak Bekerja	TB

Tabel 3.5 menjelaskan tentang macam macam pekerjaan yang dimiliki oleh orang tua siswa. Wiraswasta dikategorikan W, karyawan swasta dikategorikan KS, guru/dosen dikategorikan GD, PNS/TNI/Polri dikategorikan PTP, buruh dikategorikan sebagai B, petani dikategorikan sebagai P, nelayan dikategorikan

sebagai N, supir dikategorikan sebagai S, ibu rumah tangga dikategorikan sebagai IRT, dan tidak bekerja dikategorikan sebagai TB.

Tabel 3.6 Kategori Penghasilan Orang Tua

Kriteria Penghasilan (Rp)	Parameter	
	0 - < 500.000	Sangat Rendah
500.000 – 999.999	Rendah	R
1.000.000 – 2.999.999	Sedang	S
3.000.000 – 4.999.999	Tinggi	T
\geq 5.000.000	Sangat Tinggi	ST

Tabel 3.6 menjelaskan kategori penghasilan orang tua siswa. Orang tua yang memiliki penghasilan 0 sampai < Rp. 500.000 dikategorikan sangat rendah (SR), penghasilan Rp. 500.000 sampai Rp. 999.999 dikategorikan rendah (R), penghasilan Rp. 1.000.000 sampai Rp. 2.999.999 dikategorikan sedang (S), penghasilan Rp. 3.000.000 sampai Rp. 4.999.999 dikategorikan tinggi (T), dan orang tua yang berpenghasilan \geq Rp. 5.000.000 dikategorikan sangat tinggi (ST).

Tabel 3.7 Kategori Jumlah Tanggungan

Kriteria Jumlah Tanggungan	Parameter	
	1 – 2	Cukup
\geq 3	Banyak	B

Tabel 3.7 menjelaskan kategori jumlah anak yang menjadi tanggungan orang tua. Jumlah anak yang menjadi tanggungan orang tua dengan jumlah 1-2 orang dikategorikan cukup (C), sedangkan jumlah anak yang menjadi tanggungan orang tua dengan jumlah \geq 3 orang dikategorikan banyak (B). Gabungan data yang telah ditransformasikan dapat dilihat pada tabel 3.8

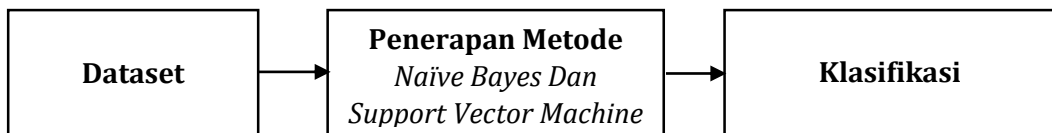
Tabel 3.8 Dataset

No	Nama	Pekerjaa n Ayah	Pekerjaa n Ibu	Penghasil an Ayah	Penghasil an Ibu	Tangg ungan	Kategori
1	Siswa 1	W	KS	T	T	B	Tepat
2	Siswa 2	W	TB	T	SR	C	Tepat
3	Siswa 3	W	IRT	S	SR	B	Tepat
4	Siswa 4	TB	KS	SR	T	C	Tepat
5	Siswa 5	KS	IRT	T	SR	C	Terlambat
...
244	Siswa 244	TB	B	SR	S	C	Terlambat
245	Siswa 245	N	TB	S	SR	C	Tepat
246	Siswa 246	KS	TB	T	SR	C	Tepat
247	Siswa 247	W	PTP	T	T	B	Terlambat
248	Siswa 248	N	TB	SR	SR	B	Tepat

3.1.2 Pelaksanaan

Setelah melewati tahap perencanaan, selanjutnya adalah tahapan pelaksanaan.

Desain sistem pada tahap pelaksanaan dapat dilihat pada gambar 3.2



Gambar 3.2 Desain Sistem

Pada gambar 3.2 merupakan desain sistem. Proses input data dilakukan dengan memasukan dataset selanjutnya dilakukan penerapan metode menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*. Setelah itu dilakukan klasifikasi keterlambatan pembayaran SPP.

Selanjutnya menghitung akurasi menggunakan *confusion matrix*. Menurut Paramitha *et al.*, (2023) *confusion matrix* menggunakan tabel matriks yang menampilkan hasil klasifikasi data *testing* berdasarkan data *training*. Berikut adalah gambar *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai actual yang berbeda.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <small>Type I Error</small>
	0 (Negative)	FN (False Negative) <small>Type II Error</small>	TN (True Negative)

Gambar 3.3 Confusion Matrix

Keterangan:

- *True Positive* (TP) adalah data positif yang diprediksi benar.
- *True Negative* (TN) adalah data negatif yang diprediksi benar.
- *False Positive* (FP) adalah data negatif tapi diprediksi sebagai data positif
- *False Negative* (FN) adalah data positif tapi diprediksi sebagai data negatif.

3.1.3 Pengamatan

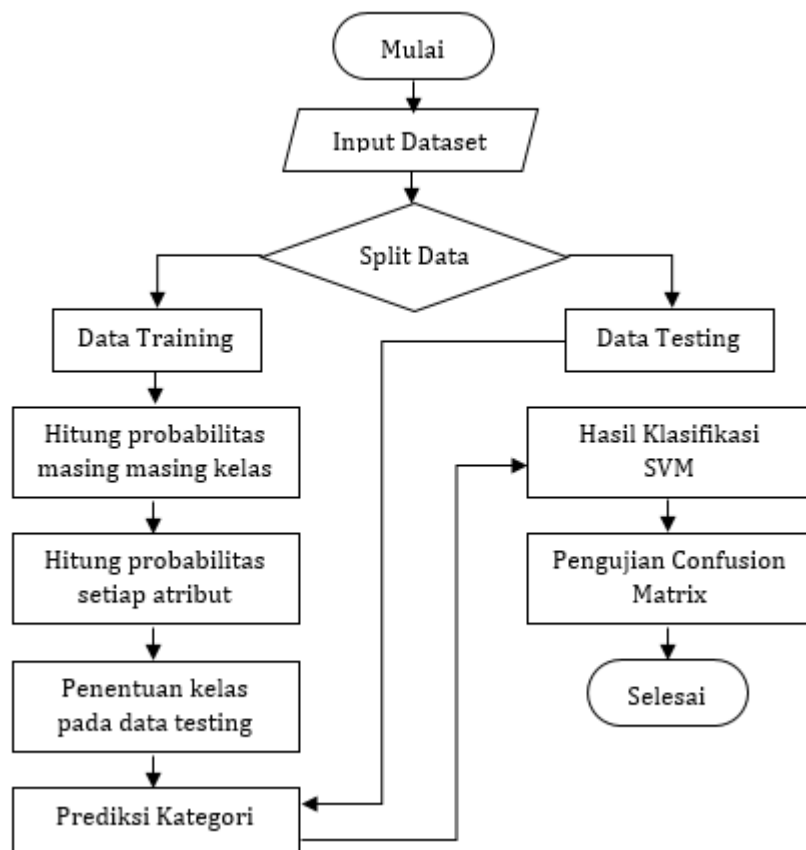
Pada tahap pengamatan dilakukan dengan melihat tingkat akurasi yang telah diperoleh kemudian melakukan evaluasi. Tahap ini bertujuan untuk mengetahui hasil dari penelitian yang telah dilakukan dan jika memungkinkan hasil dari penelitian ini dapat dikembangkan untuk penelitian selanjutnya.

BAB IV

NAÏVE BAYES

4.1 Desain

Menurut Medhat *et al.*, (2014) *Naïve Bayes* merupakan metode pengklasifikasian untuk menghitung probabilitas kelas. Desain merupakan tahapan-tahapan yang dilakukan dalam pemecahan masalah. Desain diperlukan untuk mengetahui langkah langkah yang digunakan peneliti dalam melakukan penelitian. Desain tahapan algoritma *Naïve Bayes* adalah sebagai berikut:



Gambar 4.1 Desain Tahapan Naive Bayes
(sumber: diolah dari Ginantra *et al.*, 2021)

Tahapan klasifikasi dengan algoritma *Naïve Bayes* dapat dilihat pada gambar 4.1. Persamaan teorema *Bayes* berdasarkan Muslim *et al.*, (2019) dapat dilihat pada persamaan 4.1

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)} \dots\dots\dots (4.1)$$

Keterangan:

P(H) : Probabilitas awal (prior) hipotesis H yang terjadi tanpa memandang bukti

P(E) : Probabilitas awal (prior) bukti E yang terjadi tanpa memandang bukti

P(H|E) : Probabilitas bebas bersyarat pada hipotesis H jika diberikan bukti E

P(E|H) : Probabilitas bukti E yang terjadi akan mempengaruhi hipotesis H

Langkah pertama dilakukan perhitungan probabilitas dari masing masing kelas tepat dan kelas terlambat. Data yang digunakan sebagai *training* berjumlah 20 data dan testing berjumlah 5 data. Data *training* dapat dilihat pada tabel 4.1 sedangkan data *testing* dapat dilihat pada tabel 4.7

Tabel 4.1 Data Training

NO	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah	Penghasilan Ibu	Jumlah Tanggungan	Keterangan
1	Training 1	W	KS	T	T	B	Tepat
2	Training 2	W	TB	T	SR	C	Tepat
3	Training 3	W	IRT	S	SR	B	Tepat
4	Training 4	TB	KS	SR	T	C	Tepat
...
18	Training 18	P	IRT	S	S	B	Terlambat
19	Training 19	B	TB	SR	SR	B	Terlambat
20	Training 20	KS	IRT	S	SR	B	Terlambat

Selanjutnya dilakukan perhitungan probabilitas pada masing masing kelas:

$$P(\text{Tepat}) = \frac{15}{20} = 0,750$$

$$P(\text{Terlambat}) = \frac{5}{20} = 0,250$$

Setelah mengetahui probabilitas masing masing kelas, selanjutnya menghitung probabilitas masing masing atribut. Hasil perhitungan probabilitas atribut pekerjaan ayah adalah sebagai berikut:

Tabel 4.2 Probabilitas Pekerjaan Ayah

Pekerjaan Ayah	Jumlah Kejadian		Probabilitas	
	Tepat	Terlambat	Tepat	Terlambat
B	1	1	0,067	0,200
GD	1	0	0,067	0,000
KS	4	3	0,267	0,600
N	0	0	0,000	0,000
P	1	1	0,067	0,200
PTP	0	0	0,000	0,000
S	0	0	0,000	0,000
TB	1	0	0,067	0,000
W	7	0	0,467	0,000

Pada tabel 4.2 diketahui pekerjaan ayah dengan atribut buruh pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200. Atribut guru atau dosen pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut karyawan swasta pada data actual tepat sebanyak 4 *record* memperoleh nilai probabilitas sebesar 0,267 sedangkan data actual terlambat sebanyak 3 *record* memperoleh nilai probabilitas sebesar 0,600.

Selanjutnya atribut nelayan pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut petani pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200. Atribut PNS/TNI/Polri pada data actual tepat sebanyak 0 *record*

memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut supir pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000.

Selanjutnya atribut tidak bekerja pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut wiraswasta pada data actual tepat sebanyak 7 *record* memperoleh nilai probabilitas sebesar 0,467 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Selanjutnya adalah menghitung probabilitas atribut pekerjaan ibu. Hasil perhitungan probabilitas atribut pekerjaan ibu adalah sebagai berikut:

Tabel 4.3 Probabilitas Pekerjaan Ibu

Pekerjaan Ibu	Jumlah Kejadian		Probabilitas	
	Tepat	Terlambat	Tepat	Terlambat
B	0	0	0,000	0,000
BD	0	0	0,000	0,000
GD	0	0	0,000	0,000
IRT	3	3	0,200	0,600
KS	3	0	0,200	0,000
P	1	0	0,067	0,000
PG	0	0	0,000	0,000
PTP	0	0	0,000	0,000
TB	7	2	0,467	0,400
W	1	0	0,067	0,000

Pada tabel 4.3 diketahui pekerjaan ibu dengan atribut buruh pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut bidan desa pada data actual tepat sebanyak 0 *record* memperoleh

nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut guru atau dosen pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000.

Selanjutnya atribut ibu rumah tangga pada data actual tepat sebanyak 3 *record* memperoleh nilai probabilitas sebesar 0,200 sedangkan data actual terlambat sebanyak 3 *record* memperoleh nilai probabilitas sebesar 0,600. Atribut karyawan swasta pada data actual tepat sebanyak 3 *record* memperoleh nilai probabilitas sebesar 0,200 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut petani pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut pedagang pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000.

Selanjutnya atribut PNS/TNI/Polri pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut tidak bekerja pada data actual tepat sebanyak 7 *record* memperoleh nilai probabilitas sebesar 0,467 sedangkan data actual terlambat sebanyak 2 *record* memperoleh nilai probabilitas sebesar 0,400. Atribut wiraswasta pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual

terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Selanjutnya adalah menghitung probabilitas atribut penghasilan ayah. Hasil perhitungan probabilitas atribut penghasilan ayah adalah sebagai berikut:

Tabel 4.4 Probabilitas Penghasilan Ayah

Penghasilan Ayah	Jumlah Kejadian		Probabilitas	
	Tepat	Terlambat	Tepat	Terlambat
SR	1	1	0,067	0,200
R	1	0	0,067	0,000
S	4	2	0,267	0,400
T	7	1	0,467	0,200
ST	2	1	0,133	0,200

Pada tabel 4.4 diketahui penghasilan ayah dengan atribut sangat rendah pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067. Atribut rendah pada data actual tepat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,067 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut sedang pada data actual tepat sebanyak 4 *record* memperoleh nilai probabilitas sebesar 0,267 sedangkan data actual terlambat sebanyak 2 *record* memperoleh nilai probabilitas sebesar 0,400.

Selanjutnya atribut tinggi pada data actual tepat sebanyak 7 *record* memperoleh nilai probabilitas sebesar 0,467 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200. Atribut sangat tinggi pada data actual tepat sebanyak 2 *record* memperoleh nilai probabilitas sebesar 0,133 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200. Selanjutnya adalah menghitung probabilitas atribut penghasilan ibu. Hasil perhitungan probabilitas atribut penghasilan ibu adalah sebagai berikut:

Tabel 4.5 Probabilitas Penghasilan Ibu

Penghasilan Ibu	Jumlah Kejadian		Probabilitas	
	Tepat	Terlambat	Tepat	Terlambat
SR	9	4	0,600	0,800
R	0	0	0,000	0,000
S	2	1	0,133	0,200
T	4	0	0,267	0,000
ST	0	0	0,000	0,000

Pada tabel 4.5 diketahui penghasilan ibu dengan atribut sangat rendah pada data actual tepat sebanyak 9 *record* memperoleh nilai probabilitas sebesar 0,600 sedangkan data actual terlambat sebanyak 4 *record* memperoleh nilai probabilitas sebesar 0,800. Atribut rendah pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut sedang pada data actual tepat sebanyak 2 *record* memperoleh nilai probabilitas sebesar 0,133 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200.

Selanjutnya atribut tinggi pada data actual tepat sebanyak 4 *record* memperoleh nilai probabilitas sebesar 0,267 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Atribut sangat tinggi pada data actual tepat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000 sedangkan data actual terlambat sebanyak 0 *record* memperoleh nilai probabilitas sebesar 0,000. Selanjutnya adalah menghitung probabilitas atribut jumlah tanggungan. Hasil perhitungan probabilitas atribut jumlah tanggungan adalah sebagai berikut:

Tabel 4.6 Probabilitas Jumlah Tanggungan

Jumlah Tanggungan	Jumlah Kejadian		Probabilitas	
	Tepat	Terlambat	Tepat	Terlambat
C	6	1	0,400	0,200
B	9	4	0,600	0,800

Pada tabel 4.6 diketahui jumlah tanggungan dengan atribut cukup pada data actual tepat sebanyak 6 *record* memperoleh nilai probabilitas sebesar 0,400 sedangkan data actual terlambat sebanyak 1 *record* memperoleh nilai probabilitas sebesar 0,200. Atribut banyak pada data actual tepat sebanyak 9 *record* memperoleh nilai probabilitas sebesar 0,600 sedangkan data actual terlambat sebanyak 4 *record* memperoleh nilai probabilitas sebesar 0,800.

Tabel 4.7 Data Testing

NO	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah	Penghasilan Ibu	Jumlah Tanggungan	Klasifikasi
1	Testing 1	KS	IRT	S	SR	B	?
2	Testing 2	KS	TB	T	SR	C	?
3	Testing 3	KS	P	T	S	B	?
4	Testing 4	KS	IRT	T	SR	C	?
5	Testing 5	KS	TB	ST	SR	B	?

Tabel 4.7 merupakan data yang digunakan sebagai *testing*. Setelah mengetahui probabilitas dari masing masing atribut, langkah selanjutnya adalah melakukan pengujian hasil klasifikasi menggunakan data *testing*. Berikut adalah perhitungan probabilitas data tersebut masuk dalam kategori tepat atau terlambat.

$$\begin{aligned}
 P(\text{Testing 1}|\text{Tepat}) &= P(\text{KS}|\text{Tepat}) * P(\text{IRT}|\text{Tepat}) * P(\text{S}|\text{Tepat}) * P(\text{SR}|\text{Tepat}) \\
 &\quad * P(\text{B}|\text{Tepat}) * P(\text{Tepat}) \\
 &= 0,267 * 0,200 * 0,267 * 0,600 * 0,600 * 0,750 \\
 &= 0,00384
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Testing 1}|\text{Terlambat}) &= P(\text{KS}|\text{Terlambat}) * P(\text{IRT}|\text{Terlambat}) * P(\text{S}|\text{Terlambat}) \\
 &\quad * P(\text{SR}|\text{Terlambat}) * P(\text{B}|\text{Terlambat}) * P(\text{Terlambat}) \\
 &= 0,600 * 0,600 * 0,400 * 0,800 * 0,800 * 0,250 \\
 &= 0,02304
 \end{aligned}$$

Berdasarkan perhitungan diperoleh probabilitas data *testing* 1 masuk kategori tepat (0,00384) lebih kecil dari probabilitas siswa baru masuk kategori terlambat (0,02304) sehingga data testing 1 dapat diperkirakan akan masuk kategori terlambat. Selanjutnya melakukan perhitungan untuk data lainnya dengan

menggunakan tahap perhitungan yang sama dengan data testing 1. Berikut adalah hasil klasifikasi dari data *testing*:

Tabel 4.8 Hasil Klasifikasi

NO	Nama	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Ayah	Penghasilan Ibu	Jml Tanggungan	Probabilitas Tepat	Probabilitas Terlambat	Hasil Klasifikasi
1	Testing 1	KS	IRT	SR	S	B	0,00384	0,02304	Terlambat
2	Testing 2	KS	TB	SR	T	C	0,01045	0,00192	Tepat
3	Testing 3	KS	P	S	T	B	0,00050	0,00000	Tepat
4	Testing 4	KS	IRT	SR	T	C	0,00448	0,00288	Tepat
5	Testing 5	KS	TB	SR	ST	B	0,00448	0,00768	Terlambat

Berdasarkan hasil klasifikasi dapat diketahui bahwa data testing 1 diperkirakan terlambat membayar dengan memperoleh nilai probabilitas tepat lebih kecil dari probabilitas terlambat. Selanjutnya data testing 2, 3 dan 4 diperkirakan tepat membayar dengan memperoleh nilai probabilitas tepat lebih besar dari probabilitas terlambat. Selanjutnya data testing 5 diperkirakan tepat membayar dengan memperoleh nilai probabilitas tepat lebih kecil dari probabilitas terlambat.

4.2 Uji Coba

Tahap ini dilakukan uji coba penerapan algoritma *Naïve Bayes* menggunakan 5 percobaan. Percobaan pertama dilakukan menggunakan split data 90:10 dengan 224 (*training*) dan 24 (*testing*). Percobaan kedua dilakukan menggunakan split data 80:20 dengan 198 (*training*) dan 50 (*testing*). Percobaan ketiga dilakukan menggunakan split data 70:30 dengan 174 (*training*) dan 74 (*testing*). Percobaan keempat dilakukan menggunakan split data 60:40 dengan 148 (*training*) dan 100 (*testing*). Percobaan kelima dilakukan menggunakan split data 50:50 dengan 124 (*training*) dan 124 (*testing*). Selanjutnya pada masing masing percobaan akan dilakukan penilaian performa model menggunakan *confusion matrix*.

4.2.1 Percobaan Pertama (Split Data 90:10)

Percobaan pertama dilakukan menggunakan pembagian data 90:10 dengan 224 data *training* dan 24 data *testing*. Hasil uji coba pada percobaan pertama adalah sebagai berikut:

Tabel 4.9 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	103	61
2	Pred.Terlambat	27	33

Tabel 4.9 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 103 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 27 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 61 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 33 *record* (TN).

Selanjutnya adalah mengukur performa model berdasarkan tingkat *accuracy*, *precision*, *recall* dan *f1-score*. *Accuracy* dapat diartikan sebagai tingkat kedekatan antara nilai prediksi dan nilai actual. Rumus *accuracy* menurut Indrayuni (2018) dapat dilihat pada persamaan 4.2

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \quad \dots\dots\dots (4.2) \\
 &= \frac{33 + 103}{33 + 27 + 103 + 61} \times 100\% \\
 &= \frac{136}{224} \times 100\% \\
 &= 60,71\%
 \end{aligned}$$

Precision adalah kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan. *Precision* merupakan perbandingan antara *True Positive* (TP)

dengan banyaknya data yang diprediksi positif. Rumus *precision* menurut Paramitha *et al.*, (2023) dapat dilihat pada persamaan 4.3

$$\begin{aligned}
 \textit{Precision} &= \frac{TP}{TP + FP} \times 100\% && \dots\dots\dots (4.3) \\
 &= \frac{103}{103 + 61} \times 100\% \\
 &= \frac{103}{164} \times 100\% \\
 &= 62,80\%
 \end{aligned}$$

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Recall* secara definisi adalah perbandingan antara *True Positive* (TP) dengan banyaknya data yang sebenarnya positif. Rumus *recall* menurut Ginting *et al.*, (2020) dapat dilihat pada persamaan 4.4

$$\begin{aligned}
 \textit{Recall} &= \frac{TP}{TP + FN} \times 100\% && \dots\dots\dots (4.4) \\
 &= \frac{103}{103 + 27} \times 100\% \\
 &= \frac{103}{130} \times 100\% \\
 &= 79,23\%
 \end{aligned}$$

f1-score adalah nilai rata rata harmonic dari *precision* dan *recall*. Rumus *f1-score* menurut Paramitha *et al.*, (2023) dapat dilihat pada persamaan 4.5

$$\begin{aligned}
 \textit{f1-score} &= 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} && \dots\dots\dots (4.4) \\
 &= 2 \times \frac{62,80 \times 79,23}{62,80 + 79,23} \\
 &= 2 \times \frac{4,975,644}{142,03} \\
 &= 70,07\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Naïve Bayes* pada percobaan pertama diperoleh nilai *accuracy* sebesar 60.71%, nilai *precision* sebesar 62,80%, nilai *recall* sebesar 79,23% dan nilai *f1-score* sebesar 70,07%.

4.2.2 Percobaan Kedua (Split Data 80:20)

Percobaan kedua dilakukan menggunakan pembagian data 80:20 dengan 198 data *training* dan 50 data *testing*. Hasil uji coba percobaan kedua adalah sebagai berikut:

Tabel 4.10 Uji Coba Percobaan Kedua

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	88	40
2	Pred.Terlambat	27	43

Tabel 4.10 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 88 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 27 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 40 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 43 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{43 + 88}{43 + 27 + 88 + 40} \times 100\% \\
 &= \frac{131}{198} \times 100\% \\
 &= 66,16\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{88}{88 + 40} \times 100\%
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{88}{128} \times 100\% \\
 &= 68,75\% \\
 \\
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{88}{88 + 27} \times 100\% \\
 &= \frac{88}{115} \times 100\% \\
 &= 76,52\% \\
 \\
 \text{f1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{68,75 \times 76,52}{68,75 + 76,52} \\
 &= 2 \times \frac{5.260,75}{145,27} \\
 &= 72,43\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Naïve Bayes* pada percobaan kedua diperoleh nilai *accuracy* sebesar 66,16%, nilai *precisión* sebesar 68,75%, nilai *recall* sebesar 76,52% dan nilai *f1-score* sebesar 72,43%.

4.2.3 Percobaan Ketiga (Split Data 70:30)

Percobaan ketiga dilakukan menggunakan pembagian data 70:30 dengan 174 data *training* dan 74 data *testing*. Hasil uji coba percobaan ketiga adalah sebagai berikut:

Tabel 4.11 Uji Coba Percobaan Ketiga

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	75	40
2	Pred.Terlambat	26	33

Tabel 4.11 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 75 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat

sebanyak 26 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 40 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 33 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{33 + 75}{33 + 26 + 75 + 40} \times 100\% \\
 &= \frac{108}{174} \times 100\% \\
 &= 62,07\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{75}{75 + 40} \times 100\% \\
 &= \frac{75}{115} \times 100\% \\
 &= 65,22\%
 \end{aligned}$$

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{75}{75 + 26} \times 100\% \\
 &= \frac{75}{101} \times 100\% \\
 &= 74,26\%
 \end{aligned}$$

$$\begin{aligned}
 f1-score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 &= 2 \times \frac{65,22 \times 74,26}{65,22 + 74,26} \\
 &= 2 \times \frac{4.843,24}{139,48} \\
 &= 69,45\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Naïve Bayes* pada percobaan ketiga diperoleh nilai *accuracy* sebesar 62,07%, nilai *precisión* sebesar 65,22%, nilai *recall* sebesar 74,26% dan nilai *f1-score* sebesar 69,45%.

4.2.4 Percobaan Keempat (Split Data 60:40)

Percobaan keempat dilakukan menggunakan pembagian data 60:40 dengan 148 data *training* dan 100 data *testing*. Hasil uji coba percobaan keempat adalah sebagai berikut:

Tabel 4.12 Uji Coba Percobaan Keempat

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	73	40
2	Pred.Terlambat	13	22

Tabel 4.12 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 73 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 13 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 40 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 22 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{22 + 73}{22 + 13 + 73 + 40} \times 100\% \\
 &= \frac{95}{148} \times 100\% \\
 &= 64,19\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{73}{73 + 40} \times 100\% \\
 &= \frac{73}{113} \times 100\% \\
 &= 64,60\%
 \end{aligned}$$

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \times 100\%
 \end{aligned}$$

$$\begin{aligned}
&= \frac{73}{73 + 13} \times 100\% \\
&= \frac{73}{86} \times 100\% \\
&= 84,88\% \\
f1\text{-score} &= 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} \\
&= 2 \times \frac{64,60 \times 84,88}{64,60 + 84,88} \\
&= 2 \times \frac{5.483,25}{149,48} \\
&= 73,37\%
\end{aligned}$$

Berdasarkan hasil pengukuran performa model *Naïve Bayes* pada percobaan keempat diperoleh nilai *accuracy* sebesar 64,19%, nilai *precisión* sebesar 64,60%, nilai *recall* sebesar 84,88% dan nilai *f1-score* sebesar 73,37%.

4.2.5 Percobaan Kelima (Split Data 50:50)

Percobaan kelima dilakukan menggunakan pembagian data 50:50 dengan 124 data *training* dan 124 data *testing*. Hasil uji coba percobaan kelima adalah sebagai berikut:

Tabel 4.13 Uji Coba Percobaan Kelima

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	52	28
2	Pred.Terlambat	20	24

Tabel 4.13 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 52 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 20 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 28 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 24 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 \textit{Accuracy} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{24 + 52}{24 + 20 + 52 + 28} \times 100\% \\
 &= \frac{76}{124} \times 100\% \\
 &= 61,29\%
 \end{aligned}$$

$$\begin{aligned}
 \textit{Precision} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{52}{52 + 28} \times 100\% \\
 &= \frac{52}{80} \times 100\% \\
 &= 65,00\%
 \end{aligned}$$

$$\begin{aligned}
 \textit{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{52}{52 + 20} \times 100\% \\
 &= \frac{52}{72} \times 100\% \\
 &= 72,22\%
 \end{aligned}$$

$$\begin{aligned}
 \textit{f1-score} &= 2 \times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} \\
 &= 2 \times \frac{65,00 \times 72,22}{65,00 + 72,22} \\
 &= 2 \times \frac{4.694,30}{137,22} \\
 &= 68,42\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Naïve Bayes* pada percobaan kelima diperoleh nilai *accuracy* sebesar 61,29%, nilai *precisión* sebesar 65,00%, nilai *recall* sebesar 72,22% dan nilai *f1-score* sebesar 68,42%.

4.3 Performa Naïve Bayes

Berdasarkan hasil uji coba menggunakan 5 percobaan maka rata rata performa *Naïve Bayes* adalah sebagai berikut:

Tabel 4.14 Performa Naive Bayes

NO	PERCOBAAN	DATA	PERFORMA NAÏVE BAYES			
			<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-Score</i>
1	Percobaan Pertama	90:10	60,71%	62,80%	79,23%	70,07%
2	Percobaan Kedua	80:20	66,16%	68,75%	76,52%	72,43%
3	Percobaan Ketiga	70:30	62,07%	65,22%	74,26%	69,45%
4	Percobaan Keempat	60:40	64,19%	64,60%	84,88%	73,37%
5	Percobaan Kelima	50:50	61,29%	65,00%	72,22%	68,42%
Rata Rata			62,88%	65,27%	77,42%	70,75%

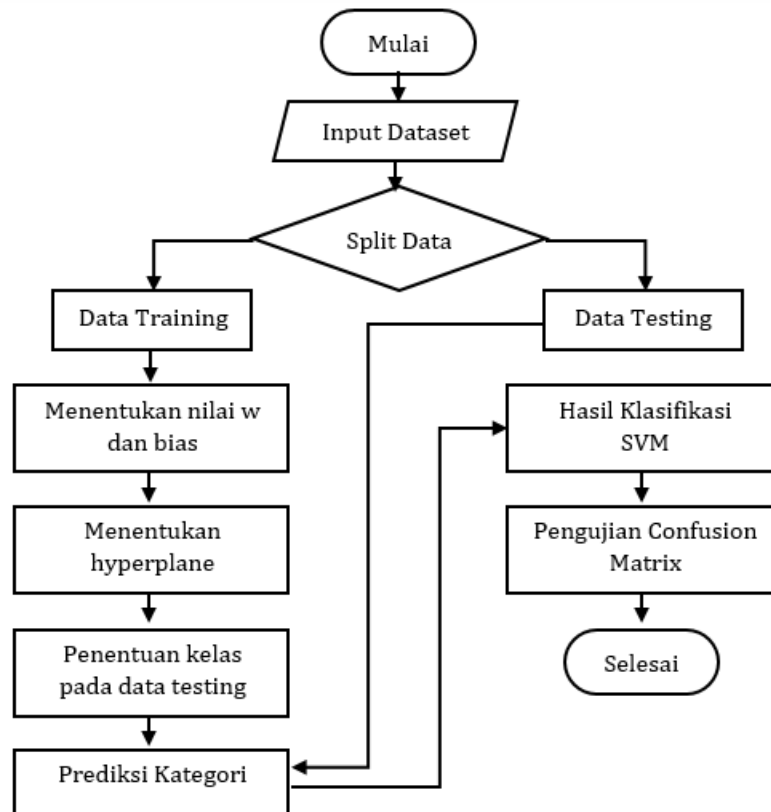
Pada tabel 4.14 diketahui nilai *accuracy* dan *precision* tertinggi diperoleh pada percobaan kedua yaitu 66,16% dan 68,75%, sedangkan nilai *recall* dan *f1-score* tertinggi diperoleh pada percobaan keempat yaitu 84,88% dan 73,37%. Rata rata performa *Naïve Bayes* adalah *accuracy* sebesar 62,88%, nilai *precision* 65,27%, nilai *recall* 77,42% dan nilai *f1-score* 70,75%.

BAB V

SUPPORT VECTOR MACHINE

5.1 Desain

Menurut Noble (2006) *Support Vector Machine* (SVM) adalah algoritma komputer yang belajar melalui contoh untuk menetapkan label pada objek. Desain tahapan merupakan tahapan-tahapan yang dilakukan dalam pemecahan masalah. Desain tahapan diperlukan untuk mengetahui langkah langkah yang digunakan peneliti dalam melakukan penelitian. Desain tahapan SVM adalah sebagai berikut:



Gambar 5.1 Desain Alur SVM
(sumber: diolah dari Muslim *et al.*, 2019)

Pada gambar 5.1 langkah pertama adalah input dataset ke dalam sistem, selanjutnya data tersebut akan dilakukan pembagian data *training* dan *testing*. Setelah itu dilakukan penentuan nilai vector bobot (w) dan bias (b). Pada dataset

memiliki 5 fitur maka w akan memiliki 5 fitur juga ($w_1, w_2, w_3, w_4,$ dan w_5) dengan formula yang akan digunakan untuk meminimalkan margin adalah sebagai berikut:

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + \dots + w_5^2)$$

dengan $y_i (w_i x_i + b) > 1, i= 1,2,3,4,5,\dots,n$

$y_i(w_1.x_1 + w_2.x_2 + w_3.x_3 + w_4.x_4 + w_5.x_5) \geq +1$ untuk kategori tepat)

$y_i(w_1.x_1 + w_2.x_2 + w_3.x_3 + w_4.x_4 + w_5.x_5) \geq -1$ untuk kategori terlambat)

Pada proses klasifikasi jika data pada kelas tepat akan diberi label +1 sedangkan data pada kelas terlambat diberi label -1. Selanjutnya adalah menentukan *hyperplane* pada masing masing kelas setelah itu penetapan kelas berdasarkan nilai terbesar dari *hyperplane*. Setelah itu dilakukan uji coba penerapan model *Support Vector Machine* melalui 5 percobaan.

5.2 Uji Coba

Tahap ini dilakukan uji coba penerapan algoritma *Support Vector Machine* menggunakan 5 percobaan. Percobaan pertama dilakukan menggunakan split data 90:10 dengan 224 (*training*) dan 24 (*testing*). Percobaan kedua dilakukan menggunakan split data 80:20 dengan 198 (*training*) dan 50 (*testing*). Percobaan ketiga dilakukan menggunakan split data 70:30 dengan 174 (*training*) dan 74 (*testing*). Percobaan keempat dilakukan menggunakan split data 60:40 dengan 148 (*training*) dan 100 (*testing*). Percobaan kelima dilakukan menggunakan split data 50:50 dengan 124 (*training*) dan 124 (*testing*). Selanjutnya pada masing masing percobaan akan dilakukan penilaian performa model menggunakan *confusion matrix*.

5.2.1 Percobaan Pertama (Split Data 90:10)

Percobaan pertama dilakukan menggunakan pembagian data 90:10 dengan 224 data *training* dan 24 data *testing*. Hasil uji coba percobaan pertama adalah sebagai berikut:

Tabel 5.1 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	122	77
2	Pred.Terlambat	8	17

Tabel 5.2 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 122 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 8 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 77 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 17 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{17 + 122}{17 + 8 + 122 + 77} \times 100\% \\
 &= \frac{139}{224} \times 100\% \\
 &= 62,05\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{122}{122 + 77} \times 100\% \\
 &= \frac{122}{199} \times 100\% \\
 &= 61,31\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{122}{122 + 8} \times 100\% \\
 &= \frac{122}{130} \times 100\% \\
 &= 93,85\%
 \end{aligned}$$

$$\begin{aligned}
 f1\text{-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{61,31 \times 93,85}{61,31 + 93,85} \\
 &= 2 \times \frac{5.753,95}{155,16} \\
 &= 74,17\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan pertama diperoleh nilai *accuracy* sebesar 62,05%, nilai *precisión* sebesar 61,31%, nilai *recall* sebesar 93,85% dan nilai *f1-score* sebesar 74,17%.

5.2.2 Percobaan Kedua (Split Data 80:20)

Percobaan kedua dilakukan menggunakan pembagian data 80:20 dengan 198 data *training* dan 50 data *testing*. Hasil uji coba percobaan kedua adalah sebagai berikut:

Tabel 5.2 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	111	69
2	Pred.Terlambat	4	14

Tabel 5.2 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 111 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 4 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 69 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat

sebanyak 14 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{14 + 111}{14 + 4 + 111 + 69} \times 100\% \\
 &= \frac{125}{198} \times 100\% \\
 &= 63,13\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{111}{111 + 69} \times 100\% \\
 &= \frac{111}{180} \times 100\% \\
 &= 61,67\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{111}{111 + 4} \times 100\% \\
 &= \frac{111}{115} \times 100\% \\
 &= 96,52\%
 \end{aligned}$$

$$\begin{aligned}
 f1\text{-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{61,67 \times 96,52}{61,67 + 96,52} \\
 &= 2 \times \frac{5951,19}{158,19} \\
 &= 75,26\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan kedua diperoleh nilai *accuracy* sebesar 63,13%, nilai *precisión* sebesar 61,67%, nilai *recall* sebesar 96,52% dan nilai *f1-score* sebesar 75,26%.

5.2.3 Percobaan Ketiga (Split Data 70:30)

Percobaan ketiga dilakukan menggunakan pembagian data 70:30 dengan 174 data *training* dan 74 data *testing*. Hasil uji coba percobaan ketiga adalah sebagai berikut:

Tabel 5.3 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	92	54
2	Pred.Terlambat	9	19

Tabel 5.3 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 92 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 9 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 54 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 19 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{19 + 92}{19 + 9 + 92 + 54} \times 100\% \\
 &= \frac{111}{174} \times 100\% \\
 &= 63,79\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{92}{92 + 54} \times 100\% \\
 &= \frac{92}{146} \times 100\% \\
 &= 63,01\%
 \end{aligned}$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$\begin{aligned}
&= \frac{92}{92 + 9} \times 100\% \\
&= \frac{92}{101} \times 100\% \\
&= 91,09\%
\end{aligned}$$

$$\begin{aligned}
f1\text{-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
&= 2 \times \frac{63,01 \times 91,09}{63,01 + 91,09} \\
&= 2 \times \frac{5.739,58}{154,10} \\
&= 74,50\%
\end{aligned}$$

Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan ketiga diperoleh nilai *accuracy* sebesar 63,79%, nilai *precision* sebesar 63,01%, nilai *recall* sebesar 91,09% dan nilai *f1-score* sebesar 74,50%.

5.2.4 Percobaan Keempat (Split Data 60:40)

Percobaan pertama dilakukan menggunakan pembagian data 60:40 dengan 148 data *training* dan 100 data *testing*. Hasil uji coba percobaan keempat adalah sebagai berikut:

Tabel 5.4 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	83	49
2	Pred.Terlambat	3	13

Tabel 5.4 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 83 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 3 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 49 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 13 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{13 + 83}{13 + 3 + 83 + 49} \times 100\% \\
 &= \frac{96}{148} \times 100\% \\
 &= 64,86\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{83}{83 + 49} \times 100\% \\
 &= \frac{83}{132} \times 100\% \\
 &= 62,88\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{83}{83 + 3} \times 100\% \\
 &= \frac{83}{86} \times 100\% \\
 &= 96,51\%
 \end{aligned}$$

$$\begin{aligned}
 f1\text{-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= 2 \times \frac{62,88 \times 96,51}{62,88 + 96,51} \\
 &= 2 \times \frac{6.068,55}{159,39} \\
 &= 76,15\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan keempat diperoleh nilai *accuracy* sebesar 64,86%, nilai *precisión* sebesar 62,88%, nilai *recall* sebesar 96,51% dan nilai *f1-score* sebesar 76,15%.

5.2.5 Percobaan Kelima (Split Data 50:50)

Percobaan kelima dilakukan menggunakan pembagian data 50:50 dengan 124 data *training* dan 124 data *testing*. Hasil uji coba percobaan kelima adalah sebagai berikut:

Tabel 5.5 Uji Coba Percobaan Pertama

No	<i>Predicted Values</i>	<i>Actual Values</i>	
		True Tepat	True Terlambat
1	Pred.Tepat	68	41
2	Pred.Terlambat	4	11

Tabel 5.5 menunjukkan jumlah data tepat yang diklasifikasikan tepat sebanyak 68 *record* (TP) dan jumlah data tepat yang diklasifikasikan terlambat sebanyak 4 *record* (FN). Berikutnya jumlah data terlambat yang diklasifikasikan tepat sebanyak 41 *record* (FP) dan jumlah data terlambat yang diklasifikasikan terlambat sebanyak 11 *record* (TN). Selanjutnya adalah mengukur performa model sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \\
 &= \frac{11 + 68}{11 + 4 + 68 + 41} \times 100\% \\
 &= \frac{79}{124} \times 100\% \\
 &= 63,71\%
 \end{aligned}$$

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{68}{68 + 41} \times 100\% \\
 &= \frac{68}{109} \times 100\% \\
 &= 62,39\%
 \end{aligned}$$

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{68}{68 + 4} \times 100\% \\
 &= \frac{68}{72} \times 100\% \\
 &= 94,44\%
 \end{aligned}$$

$$f1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$\begin{aligned}
 &= 2 \times \frac{62,39 \times 94,44}{62,39 + 94,44} \\
 &= 2 \times \frac{5.892,11}{156,83} \\
 &= 75,14\%
 \end{aligned}$$

Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan kelima diperoleh nilai *accuracy* sebesar 62,39%, nilai *precision* sebesar 62,39%, nilai *recall* sebesar 94,44% dan nilai *f1-score* sebesar 75,14%.

5.3 Performa *Support Vector Machine*

Berdasarkan hasil uji coba menggunakan 5 percobaan maka rata rata performa *Support Vector Machine* adalah sebagai berikut:

Tabel 5.6 Performa Support Vector Machine

NO	PERCOBAAN	DATA	PERFORMA SUPPORT VECTOR MACHINE			
			<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>f1-Score</i>
1	Percobaan Pertama	90:10	62,05%	61,31%	93,85%	74,17%
2	Percobaan Kedua	80:20	63,13%	61,67%	96,52%	75,26%
3	Percobaan Ketiga	70:30	63,79%	63,01%	91,09%	74,50%
4	Percobaan Keempat	60:40	64,86%	62,88%	96,51%	76,15%
5	Percobaan Kelima	50:50	63,71%	62,39%	94,44%	75,14%
Rata Rata			63,51%	62,25%	94,48%	75,04%

Pada tabel 5.6 diketahui nilai *accuracy* tertinggi diperoleh pada percobaan keempat yaitu 64,86%, nilai *precision* tertinggi diperoleh pada percobaan ketiga yaitu 63,01%, nilai *recall* tertinggi diperoleh dari percobaan kedua yaitu 96,52% dan *f1-score* tertinggi diperoleh pada percobaan keempat yaitu 76,15%. Rata rata performa *Support Vector Machine* adalah nilai *accuracy* 63,51%, nilai *precision* 62,25%, nilai *recall* 94,48% dan nilai *f1-score* 75,04%.

BAB VI

PEMBAHASAN

Penelitian ini dilatarbelakangi dari permasalahan keterlambatan pembayaran SPP yang terjadi di MTs Lukmanul Hakim Samarinda. Pada penelitian ini menggunakan data siswa dari tahun ajaran 2021/2022 sampai 2023/2024. Berdasarkan data yang bersumber dari bendahara sekolah diketahui bahwa siswa yang terlambat membayar SPP sekitar 42% dari total seluruh siswa yang berjumlah 248 siswa. Keterlambatan pembayaran SPP menjadi masalah karena dapat berpengaruh pada proses penggajian guru, karyawan dan program peningkatan mutu pendidikan di lingkungan sekolah. Oleh karena itu penelitian ini diharapkan mampu menjadi alternatif dalam mengatasi permasalahan keterlambatan pembayaran SPP di lingkungan sekolah. Pada penelitian ini menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*.

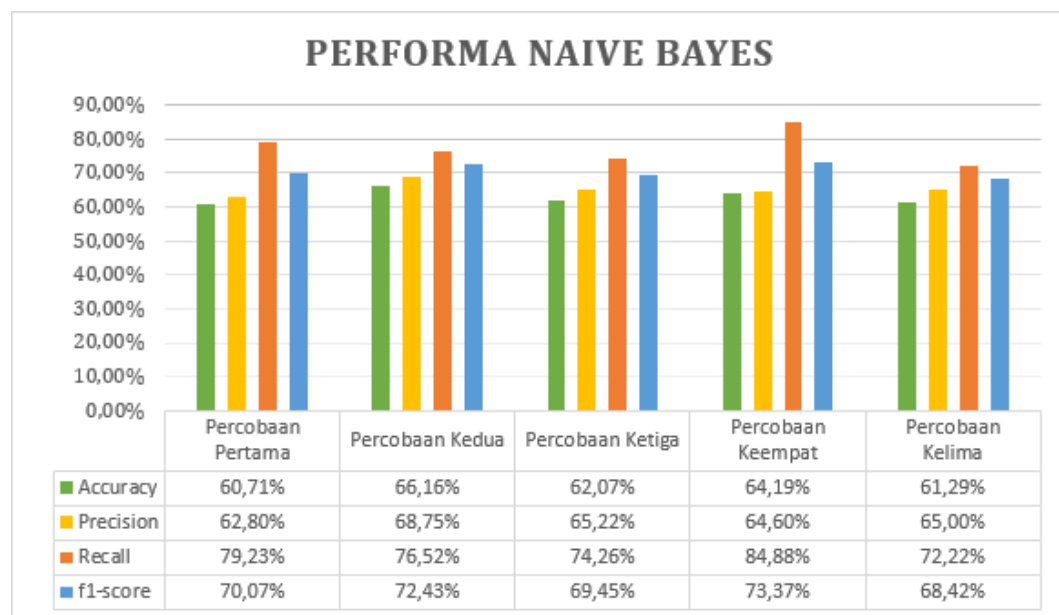
Tahapan penelitian ini diawali dari tahap perencanaan yaitu mengidentifikasi masalah yang akan diteliti, selanjutnya adalah mengumpulkan data, setelah data diperoleh selanjutnya dilakukan *pre processing* data. Tahap selanjutnya adalah tahap pelaksanaan. Pada tahap ini akan dilakukan klasifikasi menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*. Selanjutnya melakukan perhitungan menggunakan *confusion matrix* yang akan menampilkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Tahap terakhir ialah melakukan evaluasi dan penarikan kesimpulan terhadap hasil penelitian.

Uji coba penerapan algoritma *Naive Bayes* dan *Support Vector Machine* dilakukan melalui 5 percobaan. Percobaan pertama menggunakan 224 data

training dan 24 data *testing* (90:10). Percobaan kedua menggunakan 198 data *training* dan 50 data *testing* (80:20). Percobaan ketiga menggunakan 174 data *training* dan 74 data *testing* (70:30). Percobaan keempat menggunakan 148 data *training* dan 100 data *testing* (60:40). Percobaan kelima menggunakan 124 data *training* dan 124 data *testing* (50:50).

6.1. Hasil Pengujian NB dan SVM

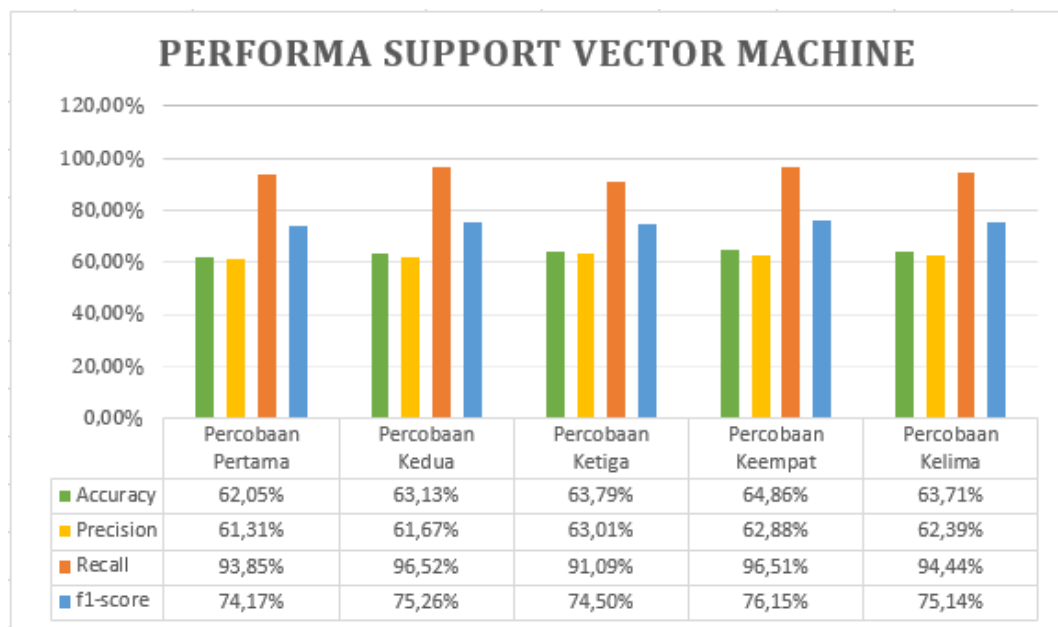
Setelah melakukan uji coba pada percobaan pertama sampai percobaan kelima diperoleh performa *Naive Bayes* dan *Support Vector Machine* berdasarkan nilai *accuracy*, nilai *precision*, nilai *recall* dan nilai *f1-score*. Performa algoritma *Naive Bayes* dan *Support Vector Machine* adalah sebagai berikut:



Gambar 6.1 Performa Naive Bayes

Pada gambar 6.1 diketahui performa *Naive Bayes* pada percobaan pertama memperoleh nilai *accuracy* (60,71%), *precision* (62,80%), *recall* (79,23%) dan *f1-score* (70,07%). Pada percobaan kedua memperoleh nilai *accuracy* (66,16%), *precision* (68,75%), *recall* (76,52%) dan *f1-score* (72,43%). Pada percobaan ketiga

memperoleh nilai *accuracy* (62,07%), *precisión* (65,22%), *recall* (74,26%) dan *f1-score* (69,45%). Pada percobaan keempat memperoleh nilai *accuracy* (64,19%), *precisión* (64,60%), *recall* (84,88%) dan *f1-score* (73,37%). Pada percobaan kelima memperoleh nilai *accuracy* (61,29%), *precisión* (65,00%), *recall* (72,22%) dan *f1-score* (68,42%). Berdasarkan grafik performa *Naive Bayes* diketahui nilai *accuracy* dan *precisión* tertinggi diperoleh dari percobaan kedua sedangkan nilai *recall* dan *f1-score* tertinggi diperoleh dari percobaan keempat.



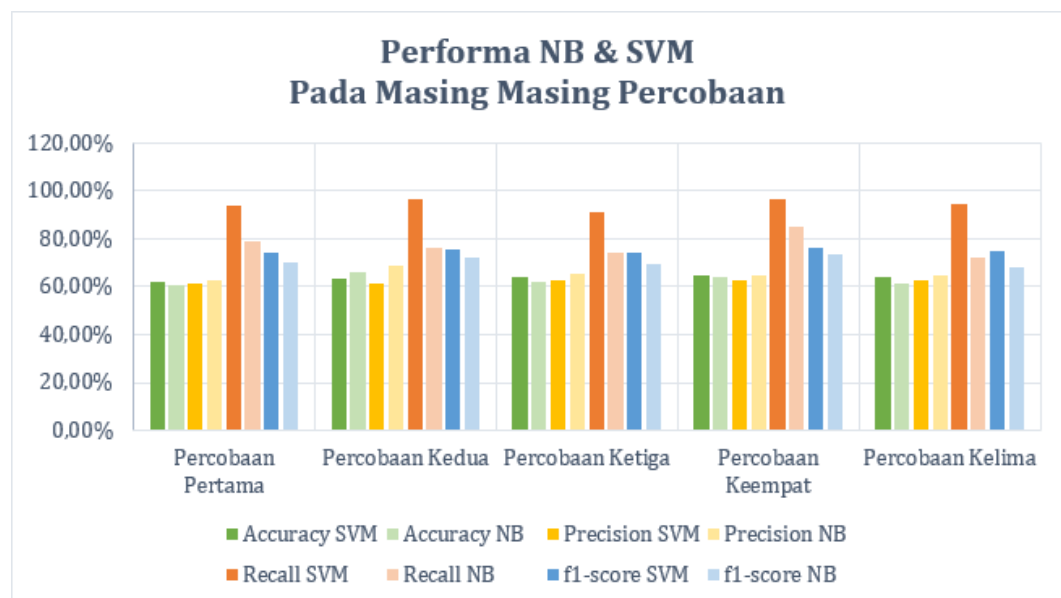
Gambar 6.2 Performa Support Vector Machine

Pada gambar 6.2 diketahui performa *Support Vector Machine* pada percobaan pertama memperoleh nilai *accuracy* (62,05%), *precisión* (61,31%), *recall* (93,85%) dan *f1-score* (74,17%). Pada percobaan kedua memperoleh nilai *accuracy* (63,13%), *precisión* (61,67%), *recall* (96,52%) dan *f1-score* (75,26%). Pada percobaan ketiga memperoleh nilai *accuracy* (63,79%), *precisión* (63,01%), *recall* (91,09%) dan *f1-score* (74,50%). Pada percobaan keempat memperoleh nilai *accuracy* (64,86%), *precisión* (62,88%), *recall* (96,51%) dan *f1-score* (76,15%).

Pada percobaan kelima memperoleh nilai *accuracy* (63,71%), *precisión* (62,39%), *recall* (94,44%) dan *f1-score* (75,14%). Berdasarkan grafik performa *Support Vector Machine* diketahui nilai *accuracy* tertinggi diperoleh dari percobaan keempat, nilai *precisión* tertinggi diperoleh dari percobaan ketiga, nilai *recall* tertinggi diperoleh dari percobaan kedua dan nilai *f1-score* tertinggi diperoleh dari percobaan keempat.

6.2. Perbandingan Performa NB dan SVM

Setelah diketahui hasil pengujian *Naïve Bayes* dan *Support Vector Machine* pada percobaan pertama sampai percobaan kelima, tahap selanjutnya adalah melakukan perbandingan performa *Naïve Bayes* dan *Support Vector Machine* berdasarkan nilai *accuracy*, nilai *precisión*, nilai *recall* dan nilai *f1-score*. Perbandingan performa *Naive Bayes* dan *Support Vector Machine* dapat dilihat pada gambar 6.3.



Gambar 6.3 Perbandingan NB dan SVM

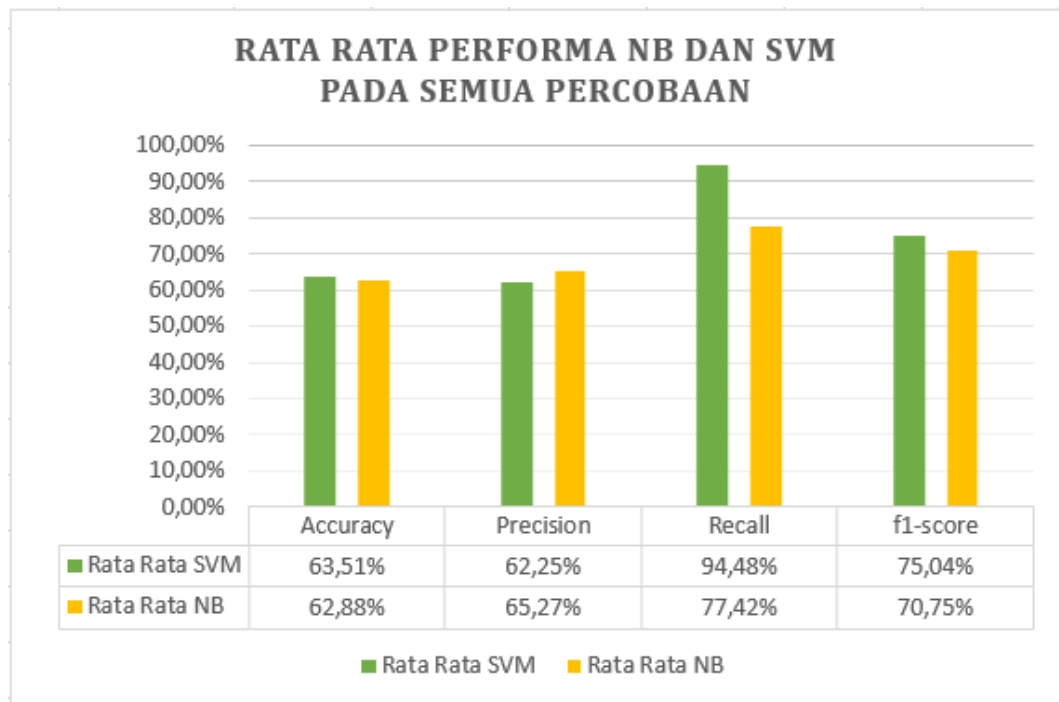
Pada gambar 6.3 diketahui pada percobaan pertama performa *Support Vector Machine* unggul pada nilai *accuracy* (62,05%), nilai *recall* (93,85%), dan nilai *f1-score* (74,17%). Sedangkan *Naïve Bayes* unggul pada nilai *precisión* (62,80%).

Pada percobaan kedua diketahui performa *Support Vector Machine* unggul pada nilai *recall* (96,52%) dan *f1-score* (75,26%). Sedangkan *Naïve Bayes* unggul pada nilai *accuracy* (66,16%) dan *precisión* (68,75%).

Pada percobaan ketiga diketahui performa *Support Vector Machine* unggul pada nilai *accuracy* (63,79%), *recall* (91,09%), dan *f1-score* (74,50%). Sedangkan *Naïve Bayes* unggul pada nilai *precisión* (65,22%).

Pada percobaan keempat diketahui performa *Support Vector Machine* unggul pada nilai *accuracy* (64,86%), *recall* (96,51%), dan *f1-score* (76,15%). Sedangkan *Naïve Bayes* unggul pada nilai *precisión* (64,60%).

Pada percobaan kelima diketahui performa *Support Vector Machine* unggul pada nilai *accuracy* (63,71%), *recall* (94,44%), dan *f1-score* (75,14%). Sedangkan *Naïve Bayes* unggul pada nilai *precisión* (65,00%). Selanjutnya rata rata performa kedua model dapat dilihat pada gambar 6.4



Gambar 6.4 Rata Rata Performa NB dan SVM

Berdasarkan perolehan rata rata performa *Naïve Bayes* dan *Support Vector Machine* pada gambar 6.4 diketahui bahwa *Support Vector Machine* memperoleh rata rata nilai *accuracy* (63,51%), *precision* (62,25%), *recall* (94,48%), dan *f1-score* (75,04%). Sedangkan *Naïve Bayes* memperoleh rata rata nilai *accuracy* (62,88%), *precision* (65,27%), *recall* (77,42%), dan *f1-score* (70,75%).

6.3. Korelasi Penelitian Dengan Nilai Nilai Islam

Penelitian klasifikasi keterlambatan SPP merupakan salah satu solusi untuk mengatasi permasalahan dalam penyelenggaraan pendidikan diantaranya adalah keterlambatan pembayaran SPP di sekolah. Klasifikasi merupakan pengelompokan data dimana data yang digunakan tersebut memiliki kelas label atau target. Konsep klasifikasi atau pengelompokan data dapat ditemukan dalam Al Qur'an surah Al An'am ayat 142 yaitu:

وَمِنَ الْأَنْعَامِ حَمُولَةً وَفَرْشًا كُلُوا مِمَّا رَزَقَكُمُ اللَّهُ وَلَا تَتَّبِعُوا خُطُوَاتِ الشَّيْطَانِ إِنَّهُ لَكُمْ عَدُوٌّ مُّبِينٌ

artinya: “dan di antara hewan-hewan ternak itu ada yang dijadikan pengangkut beban dan ada (pula) yang untuk disembelih. Makanlah rezeki yang diberikan Allah kepadamu dan janganlah kamu mengikuti langkah-langkah setan. Sesungguhnya setan itu musuh yang nyata bagimu.”

Menurut tafsir Ibnu Katsir yang diterjemahkan oleh Ghoffar E.M. *et al.*, (2004) diceritakan bahwa Allah menjadikan untuk manusia binatang ternak yang sebagian darinya dapat dijadikan sebagai kendaraan angkutan, dan sebagian yang lain dijadikan sebagai hewan potong. Seperti yang dikatakan oleh As-Sauri, dari Abu Ishaq dari Abul Ahwas, dari Abdullah sehubungan dengan makna firman Allah: maksud dari *Hamuulah* adalah unta yang dijadikan sebagai kendaraan angkutan, sedangkan *Farsy* adalah unta yang masih muda. Demikian menurut riwayat Imam Hakim yang mengatakan sanad asar ini sahih, tetapi Imam Bukhari dan Muslim tidak menyetujuikannya.

Pendidikan merupakan salah satu aspek penting dalam kehidupan manusia. Pendidikan memiliki keterkaitan dengan konsep *mu'amalah ma'allah* atau hubungan manusia dengan Allah ﷻ sebagai pencipta dan *mu'amalah ma'annas* atau hubungan manusia dengan sesamanya. Sebagaimana Allah ﷻ menegaskan pentingnya pendidikan dalam Al Qur'an surah Al Alaq ayat 1-5:

(5) أَفْرَأُ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (1) خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ (2) أَفْرَأُ وَرَبُّكَ الْأَكْرَمُ (3) الَّذِي عَلَّمَ بِالْقَلَمِ (4)
عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ

Artinya:

(1) Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan. (2) Dia telah menciptakan manusia dari segumpal darah. (3) Bacalah, dan Tuhanmulah Yang Mahamulia. (4) Yang mengajar (manusia) dengan pena. (5) Dia mengajarkan manusia apa yang tidak diketahuinya.

Dalam tafsir Ibnu Katsir yang diterjemahkan oleh Ghoffar E.M. *et al.*, (2004) diceritakan bahwa wahyu pertama yang disampaikan kepada Rasulullah ﷺ berupa mimpi yang benar dari tidurnya. Pada suatu hari saat Rasulullah ﷺ sedang berada di Gua Hira, beliau dikejutkan dengan datangnya malaikat Jibril membawa wahyu dan berkata “bacalah”. Lalu Rasulullah ﷺ menjawab “aku tidak dapat membaca”. Kemudian malaikat Jibril memeluknya hingga Rasulullah ﷺ merasa kepayahan. Lalu malaikat Jibril melepaskannya dan berkata lagi “bacalah” dan Rasulullah ﷺ menjawab “aku tidak dapat membaca”. Kemudian malaikat Jibril memeluknya untuk kedua kalinya dan melepaskannya sembari berkata lagi “bacalah” dan Rasulullah ﷺ pun menjawab “aku tidak dapat membaca”. Kemudian malaikat Jibril memeluknya untuk ketiga kalinya dan melepaskannya seraya membacakan surat Al Alaq ayat 1 sampai 5.

Secara eksplisit dalam kitab *Tarjamah Sunan Abi Daud* yang ditulis oleh H. B. Arifin & A. Syinqithy (1993) Rasulullah ﷺ bersabda:

وَمَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ

Artinya: *Barangsiapa yang menempuh suatu jalan untuk mencari ilmu, kecuali Allah mempermudah baginya jalan menuju surga.*

Hadis ini menggambarkan betapa pentingnya menuntut ilmu sampai Allah ﷻ memberikan jaminan kemudahan menuju tempat yang paling dinantikan oleh seluruh kalangan umat manusia yaitu surga.

Konsep *mu'amalah ma'annas* dapat ditemukan pada perkataan Ali bin Abi Thalib RA dalam kitab *Ta'lim al-Muta'allim* karangan Syaikh Az-Zarnuji yang diterjemahkan oleh Aljufri (2009). Ali bin Abi Thalib RA memberikan beberapa syarat dalam menuntut ilmu sebagai berikut:

أَلَا لَنْ تَنَالِ الْعِلْمَ إِلَّا بِسِتَّةٍ سَأْتِيكَ عَنْ مَجْمُوعِهَا بَيَانٌ ذِكَاؤٌ وَجُرُصٌ وَاصْطِبَارٌ وَبُلْغَةٌ وَإِرْشَادٌ أُسْتَاذٌ وَطُولٌ

زَمَانٌ

Artinya:

"Ketahuilah, engkau tidak akan memperoleh ilmu kecuali dengan bekal enam perkara, yaitu: cerdas, semangat, bersabar, memiliki bekal, petunjuk atau bimbingan guru, dan waktu yang lama"

Sebagai seorang muslim sepatutnya mampu memiliki 6 bekal tersebut dalam menuntut ilmu. Salah satu syarat penting dari nasehat tersebut adalah memiliki bekal. Menurut Hasanah (2019) memiliki bekal artinya sebagai seorang penuntut ilmu hendaknya mau mengeluarkan uang untuk keperluan pendidikan seperti membayar biaya pendidikan (SPP) dan berbagai kebutuhan lainnya dalam mendukung kegiatan belajar.

BAB VII

PENUTUP

7.1 Kesimpulan

Berdasarkan hasil penelitian dapat disimpulkan bahwa klasifikasi keterlambatan pembayaran SPP menggunakan *Naïve Bayes* dan *Support Vector Machine* dilakukan uji coba melalui 5 percobaan. Percobaan pertama dengan pembagian data 90:10, percobaan kedua dengan pembagian data 80:20, percobaan ketiga dengan pembagian data 70:30, percobaan keempat dengan pembagian data 60:40 dan percobaan kelima dengan pembagian data 50:50. Berdasarkan hasil perhitungan dari percobaan pertama sampai percobaan kelima diketahui rata rata nilai akurasi *Naïve Bayes* sebesar 62,88% sedangkan *Support Vector Machine* sebesar 63,51%.

7.2 Saran

Berikut adalah beberapa saran yang dapat dijadikan bahan pertimbangan untuk penelitian selanjutnya:

1. Penelitian selanjutnya disarankan untuk dapat melakukan optimalisasi algoritma agar dapat memperoleh tingkat performa yang lebih maksimal.
2. Penelitian selanjutnya disarankan untuk menggunakan dataset dengan jumlah yang lebih banyak agar dapat memperoleh hasil yang lebih maksimal.

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). *Informasi Interaktif, Vol 4, No 3*, 160–164.
- Aljufri, A. K. (2009). *Terjemah Ta'lim Muta'allim*. Surabaya: Mutiara Ilmu.
- 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>
- Arifin, H. B., & A. Syinqithy, D. (1993). *Tarjamah Sunan Abi Daud (Jilid IV)*. Semarang: CV. Asy Syifa'.
- Arwildayanto, Nina, L., & Warni, T. S. (2017). Manajemen Keuangan dan Pembiayaan Pendidikan. *Jabar: Penerbit Widya Padjadjaran*.
- Bahtiar, A., & Firmansyah, D. (2017). Klasifikasi Ketepatan Waktu Pembayaran SPP di Pondok Pesantren Al-Arifah Menggunakan Algoritma Naive Bayes. *Jurnal Ilmiah Manajemen Informatika dan Komputer, 01(01)*.
- Ghoffar E.M., M. A., Mu'thi, A., & Al-Atsari, A. I. (2004). *Tafsir Ibnu Katsir (Jilid 8)*. Bogor: Pustaka Imam Asy-Syafi'i.
- Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Separini, R. S., Ahmad, N., Ardiana, D. P. Y., Effendy, F., Hazriani, A. I., Sari, I. Y., Priatno, Z. G. C., Gustian, D., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis.
- 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)*. <https://doi.org/10.35585/inspir.v10i1.2535>
- Hasanah, A. N. (2019). *Enam Syarat Meraih Ilmu Menurut Sayyidina Ali Bin Abi Thalib*. Diakses pada 07 Maret 2024 dari <https://bincangsyariah.com/khazanah/enam-syarat-meraih-ilmu-menurut-sayyidina-ali-bin-abi-thalib/>
- Indrayuni, E. (2018). Komparasi Algoritma Naive Bayes dan Support Vector Machine Untuk Analisa Sentimen Review Film. *Jurnal Pilar Nusa Mandiri, 14(2)*, 175. <https://doi.org/10.33480/pilar.v14i2.918>

- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- 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>
- Muslim, M. A., Prasetyo, B., Mawarni, E. L. H., Herowati, A. J., Mirqotussa'adah, Rukmana, S. hardiyanti, & Nurzahputra, A. (2019). *Data Mining*. Universitas Negeri Semarang.
- Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, 24(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565>
- Noor Islahudin, A., & Ramadhani Wulandari, N. (2022). Manajemen Pembiayaan Pendidikan Perspektif Al-Quran. *Mindset: Jurnal Manajemen Pendidikan Islam*, 1–21. <https://doi.org/10.58561/mindset.v1i1.24>
- Norhalimi, M., & Siswa, T. A. Y. (2022). Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(3), 237–255. <https://doi.org/10.14421/jiska.2022.7.3.237-255>
- Nugraha, Y. R. A. (2021). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Penyesuaian Uang Kuliah Tunggal Terdampak Pandemi Covid-19. *Jurnal Teknik Informatika Kaputama (JTIK)*, Vol. 5(No. 2).
- Paramitha, N. Y., Nuryaman, A., Faisol, A., & Setiawan, E. (2023). *Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes*. 04(01).
- Prabowo, C., Hakim, L. N., Rohmat, C. L., & Rinaldi, A. R. (2021). Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran. *Jurnal Data Science dan Informatika (JDSI)*, 1(1).
- Rosiana, L. (2022). Analisis Kemungkinan Keterlambatan Pembayaran SPP Menggunakan Algoritma Support Vector Machine (Studi Kasus: Smp Perintis 2 Bandar Lampung). *Jurnal Ilmu Data*, Vol 2(No. 9).
- Wibowo, A. (2017). *Klasifikasi*. Binus University. Diakses dari <https://mti.binus.ac.id/2017/11/24/klasifikasi/>
- Widayani, W., & Harliana, H. (2021). Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains dan Informatika*, 7(1), 20–27. <https://doi.org/10.34128/jsi.v7i1.268>

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. <https://doi.org/10.1007/s10115-007-0114-2>