



Klasifikasi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Menggunakan Algoritma Naïve Bayes dan Support Vector Machine

Huzaifah Umar¹, R. Kusumawati^{1,*}, M. Imamudin¹, Moh. Ainur Rohman²

¹Fakultas Sains dan Teknologi, Magister Informatika, UIN Maulana Malik Ibrahim, Malang, Indonesia

²Fakultas Teknik, Program Studi Profesi Insinyur, Universitas Jember, Kota Jember, Indonesia

Email: ¹huzaifahumar23@gmail.com, ^{2,*}ririen.kusumawati@ti.uin-malang.ac.id, ³imamudin@ti.uin-malang.ac.id,

⁴ainurrohman869@gmail.com

Email Penulis Korespondensi: ririen.kusumawati@ti.uin-malang.ac.id

Abstrak—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 *Naïve Bayes* dan *Support Vector Machine*. Metode yang digunakan adalah Cross-Industry Standard Process for Data Mining (CRISP-DM). Uji coba metode dilakukan dengan 5 kali percobaan. Berdasarkan hasil pengujian diperoleh rata-rata performa *Naïve 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%).

Kata Kunci: Klasifikasi; SPP; *Naïve Bayes*; *Support Vector Machine*; CRISP-DM

Abstract—Payment delinquency of SPP is a commonly occurring issue in school. It affects the salary of teachers and staffs alongside school's various development program. The study aims to classify payment delinquencies using *Naïve Bayes* and *Support Vector Machine*. Research methode is Cross-Industry Standard Process for Data Mining (CRISP-DM). Method testing was carried out with 5 trials. Based on the test results, the average performance of *Naïve Bayes* is *accuracy* (62,88%), *precision* (65,27%), *recall* (77,42%) dan *f1-score* (70,75%). Meanwhile, the average performance of the *Support Vector Machine* is *accuracy* 63,51%), *precision* (62,25%), *recall* (94,48%) dan *f1-score* (75,04%).

Keywords: Classification; SPP; *Naïve Bayes*; *Support Vector Machine*; CRISP-DM

1. PENDAHULUAN

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. MTs Lukmanul Hakim merupakan salah satu lembaga pendidikan swasta yang berada 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. Pada umumnya SPP diterapkan oleh sekolah swasta karena dalam pengelolaan pendidikannya dibebankan kepada masyarakat atau kebijakan lokal sedangkan sekolah negeri biaya pengelolaan pendidikannya ditanggung oleh pemerintah.

Masalah yang sering muncul dalam hal pembayaran SPP adalah apabila siswa terlambat membayar tagihan SPP dari waktu yang telah ditentukan oleh sekolah. Hal ini menjadi masalah dikarenakan keterlambatan pembayaran SPP dapat berpengaruh pada proses penggajian guru dan karyawan maupun program peningkatan mutu pendidikan di lingkungan sekolah. Berdasarkan data yang diperoleh dari bendahara sekolah diketahui bahwa siswa yang terlambat membayar SPP pada tahun ajaran 2023/2024 sekitar 42% dari total seluruh 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.

Beberapa penelitian tentang klasifikasi pembayaran SPP telah dilakukan seperti penelitian tentang teknik klasifikasi pembayaran SPP menyebutkan biaya pendidikan sangat berperang penting dalam penyelenggaraan pendidikan di pondok pesantren. Keterlambatan dalam melakukan pembayaran SPP menjadi masalah yang serius di lingkungan pesantren sebab SPP menjadi sumber dana yang akan digunakan pesantren dalam mendukung penyelenggaraan pendidikan. Oleh karena itu diperlukan solusi untuk mengatasi masalah tersebut dengan melakukan klasifikasi pembayaran SPP berdasarkan tingkat ketepatan waktu pembayaran menggunakan algoritma *Naïve Bayes* sehingga dapat menjadi bahan evaluasi pihak pesantren. Dari model yang digunakan menghasilkan tingkat akurasi sebesar 63,64% (Prabowo et al., 2021).

Penelitian tentang klasifikasi ketepatan waktu pembayaran SPP mengatakan masalah yang sering muncul dalam pembayaran SPP adalah ketika para santri terlambat dalam melakukan pembayaran SPP dari batas waktu yang telah ditentukan. Hal tersebut menjadi masalah yang serius karena SPP menjadi salah satu faktor penting dalam meningkatkan kualitas pelayanan yang baik 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 dari penelitian ini adalah mengklasifikasikan pembayaran SPP berdasarkan ketepatan waktu pembayaran menggunakan algoritma *Naïve Bayes*. Hasil akurasi yang diperoleh sebesar 63,64% berdasarkan 45 data training dan 22 data testing (Bahtiar & Firmansyah, 2017).



Penelitian tentang optimasi metode *Naïve Bayes* dengan *feature selection* mengungkapkan bahwa siswa yang terlambat membayar SPP pada tahun ajaran 2017/2018 adalah sebesar 60% dari total seluruh siswa yang berjumlah 236 siswa. Hal ini menunjukkan bahwa permasalahan keterlambatan pembayaran SPP di lingkungan sekolah masih menjadi permasalahan yang serius. Oleh karena itu diperlukan solusi berupa sistem yang mampu memperkirakan siswa yang tepat atau terlambat dalam melakukan pembayaran SPP. Penelitian ini menerapkan model gabungan antara algoritma *Information Gain* dan *Naïve Bayes*. Hasil yang diperoleh sebesar 90% sedangkan jika hanya menggunakan *Naïve Bayes* hanya memperoleh nilai akurasi sebesar 80% (Muqorobin et al., 2019). Selanjutnya penelitian tentang prediksi keterlambatan pembayaran sumbangan pembangunan pendidikan sekolah melakukan penelitian dengan dataset yang sama dari Dari hasil penelitian didapatkan hasil tingkat akurasi sebesar 73% (Ginting et al., 2020).

Penelitian tentang optimasi seleksi fitur *Information Gain* pada algoritma *Naïve 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 *Naïve Bayes* dengan *Information Gain*. Sedangkan pada algoritma *K-Nearest Neighbor* dengan *Information Gain* hanya diperoleh tingkat akurasi sebesar 50,76% (Norhalimi & Siswa, 2022).

Penelitian tentang analisis kemungkinan keterlambatan pembayaran SPP di SMP Perintis 2 Bandar Lampung menggunakan algoritma *Support Vector Machine*. 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 (Rosiana, 2022).

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

Penelitian tentang analisis kemungkinan keterlambatan pembayaran SPP menggunakan algoritma C4.5. Hasil klasifikasi dari algoritma tersebut kemudian divalidasi dan dievaluasi 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% (Apandi et al., 2019).

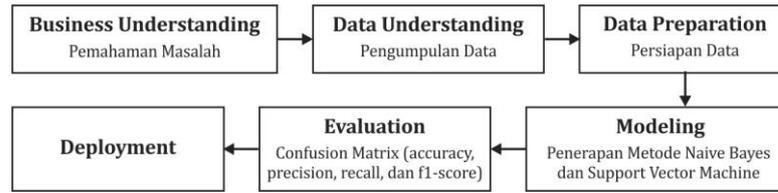
Penelitian tentang melakukan klasifikasi penyesuaian UKT mahasiswa terdampak pandemi Covid-19 menggunakan algoritma *Support Vector Machine*. Dari hasil pengujian diperoleh tingkat akurasi sebesar 83,00% dari 100 dataset yang diujicobakan menggunakan aplikasi Rapidminer (Nugraha, 2021).

Penelitian tentang klasifikasi penundaan biaya kuliah mahasiswa menggunakan *Support Vector Machine*. Penelitian ini bertujuan untuk mendapatkan nilai akurasi tertinggi menggunakan *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% (Widayani & Harliana, 2021).

Beberapa penelitian terdahulu yang telah diuraikan belum membahas tentang penerapan metode *Naïve Bayes* dan *Support Vector Machine* dalam melakukan klasifikasi keterlambatan pembayaran SPP sehingga pada penelitian kali ini akan dilakukan klasifikasi keterlambatan pembayaran SPP pada MTs Lukmanul Hakim menggunakan *Naïve Bayes* dan *Support Vector Machine*. Tujuan dari penelitian ini adalah membangun model klasifikasi keterlambatan pembayaran SPP menggunakan *Naïve Bayes* dan *Support Vector Machine*. Kedua algoritma tersebut dipilih karena merupakan bagian dari 10 algoritma terbaik dalam data mining berdasarkan paper yang dipublikasikan pada tahun 2006 oleh IEEE *International Conference on Data Mining* (Wu et al., 2008). Hasil penelitian ini diharapkan dapat dijadikan sebagai bahan pertimbangan oleh pihak sekolah dalam mengatasi keterlambatan pembayaran SPP.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang dikembangkan oleh beberapa industri yaitu *Daimler-Chrysler*, *SPSS*, dan *NCR*. CRISP-DM dipilih karena merupakan standard dalam data mining yang paling populer saat ini berdasarkan hasil polling *datascience-pm.com* (Sukesa, 2022). Tahapan CRISP-DM terdiri dari *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation dan deployment* (Shafique & Qaiser, 2014).



Gambar 1. Tahapan Penelitian

Tahapan penelitian pada gambar 1 diawali dengan tahap *business understanding* yang merupakan tahap awal dari CRISP-DM yang dilakukan untuk mendefinisikan business problem dengan sederhana dan tepat. Pada tahap *business understanding* akan diketahui masalah yang akan dipecahkan yaitu keterlambatan pembayaran SPP di MTs Lukmanul Hakim Samarinda. Selanjutnya adalah tahap *data understanding* yaitu tahap memahami struktur data. Pada *data understanding* dilakukan pengumpulan data awal. Data yang digunakan pada penelitian ini adalah data keterlambatan pembayaran SPP yang diperoleh dari bagian tata usaha dan bendahara sekolah. Data tersebut berjumlah 248 data dan memiliki atribut seperti nama pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, tanggungan dan kategori pembayaran. Selanjutnya adalah tahap *data preparation* yaitu melakukan persiapan data awal sehingga siap digunakan pada saat permodelan. Secara garis besar *data preparation* dilakukan untuk memperbaiki masalah dalam data. Kegiatan yang dilakukan antara lain seleksi data, integrasi data dan transformasi data.

Setelah melewati tahap *data preparation* selanjutnya adalah tahap modeling yaitu melakukan penerapan metode *Naïve Bayes* dan *Support Vector Machine* untuk melakukan klasifikasi keterlambatan pembayaran SPP. Setelah itu melakukan evaluasi terhadap model untuk menilai performa model menggunakan *confusion matrix*. *confusion matrix* menggunakan tabel matriks yang menampilkan hasil klasifikasi data *testing* berdasarkan data *training* (Paramitha et al., 2023). Adapun dalam tahap evaluasi akan diperoleh hasil performa dari *Naïve Bayes* dan *Support Vector Machine* berdasarkan nilai *accuracy*, *precision*, *recall* dan *f1-score*. Berikut adalah tabel *confusion matrix* dengan 4 kombinasi nilai prediksi dan nilai *actual* yang berbeda.

Tabel 1. *Confusion matrix*

<i>Predicted Values</i>	<i>Actual Values</i>	
	True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)	

Pada tabel 1 data 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 sedangkan *false negative* (FN) adalah data positif tapi diprediksi sebagai data negatif. Perhitungan performa berdasarkan model dapat dilihat pada persamaan 1, persamaan 2, persamaan 3 dan persamaan 4 berikut ini: (Indrayuni, 2018)

$$Accuracy = \frac{(TN + TP)}{(TN + FN + TP + FP)} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$f1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Selanjutnya adalah tahap *deployment* yang merupakan tahapan akhir dari metode CRISP-DM. *Deployment* bertujuan untuk mengetahui hasil dari penelitian yang telah dilakukan dan jika memungkinkan hasil dari penelitian ini dapat dikembangkan untuk penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

3.1 *Naïve Bayes*

Naïve Bayes merupakan salah satu metode klasifikasi yang bersumber pada teorema Bayes. *Naïve Bayes* merupakan metode pengklasifikasian untuk menghitung probabilitas kelas. (Medhat et al., 2014) Secara sederhana pengelompokan *Naïve Bayes* menganggap adanya suatu fitur tertentu dalam suatu kelas tidak memiliki keterikatan dengan fitur lainnya (Anam & Santoso, 2018). Persamaan teorema *Bayes* berdasarkan dapat dilihat pada persamaan 5 berikut ini (Bustami, 2014).

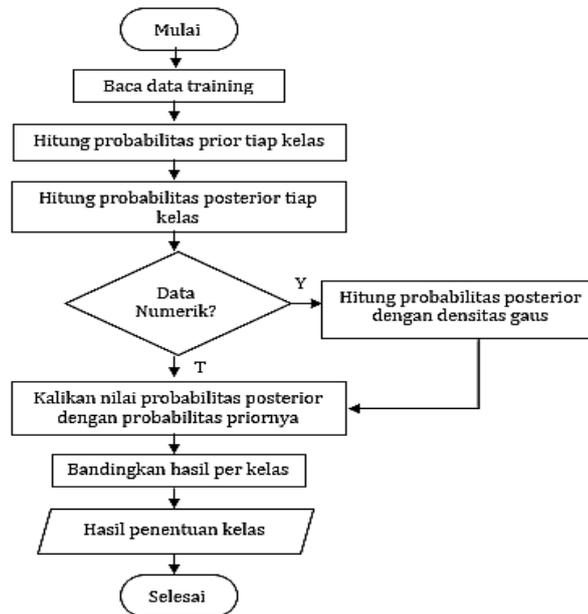
$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (5)$$

Keterangan:

X : Data dengan kelas yang belum diketahui

H : Hipotesis data X yang merupakan suatu kelas spesifik

- P(H) : Probabilitas hipotesis H (prior)
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posterior)
- P(X) : Probabilitas X
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H



Gambar 2. Alur metode *Naive Bayes*

Tahapan dalam melakukan klasifikasi menggunakan *Naive Bayes* dapat dilihat pada gambar 2. Langkah pertama dimulai dengan membaca data *training* selanjutnya dilakukan perhitungan probabilitas prior dari masing masing kelas tepat dan kelas terlambat. Setelah mengetahui probabilitas masing masing kelas, selanjutnya melakukan perhitungan probabilitas posterior setiap kelas. Selanjutnya nilai probabilitas posterior akan dikalikan dengan probabilitas priornya. Kemudian membandingkan hasil per kelas dan menghitung hasil penentuan kelas.

3.2 Uji Coba *Naive Bayes*

3.2.1 Percobaan Pertama (Data 90:10)

Percobaan pertama dilakukan menggunakan pembagian data 90:10 dengan 224 data training dan 24 data testing. Hasil klasifikasi percobaan pertama dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Uji Coba NB Percobaan Pertama

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	103	61
Pred.Terlambat	27	33

Tabel 2 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). Berdasarkan hasil pengukuran performa model *Naive 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%.

3.2.2 Percobaan Kedua (Data 80:20)

Percobaan kedua dilakukan menggunakan pembagian data 80:20 dengan 198 data training dan 50 data testing. Hasil klasifikasi percobaan kedua dapat dilihat pada tabel 3 berikut ini:

Tabel 3. Uji Coba NB Percobaan Kedua

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	88	40
Pred.Terlambat	27	43

Tabel 3 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). 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%.

3.2.3 Percobaan Ketiga (Data 70:30)

Percobaan ketiga dilakukan menggunakan pembagian data 70:30 dengan 174 data training dan 74 data testing. Hasil klasifikasi percobaan ketiga dapat dilihat pada tabel 4 berikut ini:

Tabel 4. Uji Coba NB Percobaan Ketiga

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	75	40
Pred.Terlambat	26	33

Tabel 4 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). 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%.

3.2.4 Percobaan Keempat (Data 60:40)

Percobaan keempat dilakukan menggunakan pembagian data 60:40 dengan 148 data training dan 100 data testing. Hasil klasifikasi percobaan keempat dapat dilihat pada tabel 5 berikut ini:

Tabel 5. Uji Coba NB Percobaan Keempat

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	73	40
Pred.Terlambat	13	22

Tabel 5 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). 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%.

3.2.5 Percobaan Kelima (Data 50:50)

Percobaan kelima dilakukan menggunakan pembagian data 50:50 dengan 124 data training dan 124 data testing. Hasil klasifikasi percobaan kelima dapat dilihat pada tabel 6 berikut ini:

Tabel 6. Uji Coba NB Percobaan Kelima

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	52	28
Pred.Terlambat	20	24

Tabel 6 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). 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%.

3.2.6 Performa *Naïve Bayes*

Setelah melakukan uji coba dengan 5 percobaan maka rata rata performa *Naïve Bayes* berdasarkan nilai *accuracy*, *precisión*, *recall*, dan *f1-score* dapat dilihat pada tabel 7 berikut ini:

Tabel 7. Performa *Naïve Bayes*

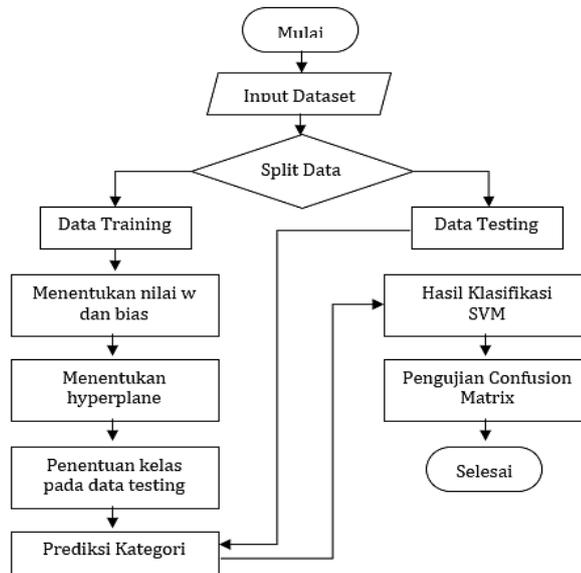
NO	PERCOBAAN	DATA	PERFORMA NAÏVE BAYES			
			Accuracy	Precision	Recall	f1-score
1	Percobaan Pertama	90:10	60,71%	62,80%	79,23%	70,07%
2	Percobaan Kedua	80:20	66,16%	68,75%	76,52%	72,43%
3	Percobaan Ketiga	70:30	62,07%	65,22%	74,26%	69,45%

NO	PERCOBAAN	DATA	PERFORMA NAÏVE BAYES			
			Accuracy	Precision	Recall	f1-score
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 7 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%, *precision* sebesar 65,27%, *recall* sebesar 77,42% dan *f1-score* sebesar 70,75%.

3.3 Support Vector Machine

Support Vector Machine (SVM) adalah algoritma komputer yang belajar melalui contoh untuk menetapkan label pada objek (Noble, 2006). *Support Vector Machine* bekerja berdasarkan *Structural Risk Minimization* (SRM) yang dirancang untuk mengolah data menjadi hyperline (pemisah). Desain alur *Support Vector Machine* dapat dilihat pada gambar 3.



Gambar 3. Alur metode *Support Vector Machine*

Pada gambar 3 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:

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + \dots + w_5^2) \tag{6}$$

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

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

$y_i (w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5) \leq -1$ 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*. Tahap selanjutnya adalah melakukan uji coba penerapan model *Support Vector Machine* melalui 5 percobaan.

3.4 Uji Coba Support Vector Machine

3.4.1 Percobaan Pertama (Data 90:10)

Percobaan pertama dilakukan menggunakan pembagian data 90:10 dengan 224 data training dan 24 data testing. Hasil klasifikasi percobaan pertama dapat dilihat pada tabel 8 berikut ini:

Tabel 8. Uji Coba SVM Percobaan Pertama

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	122	77



<i>Predicted Values</i>	<i>Actual Values</i>	
	True Tepat	True Terlambat
Pred.Terlambat	8	17

Tabel 8 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). Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan pertama diperoleh nilai *accuracy* sebesar 62,05%, nilai *precision* sebesar 61,31%, nilai *recall* sebesar 93,85% dan nilai *f1-score* sebesar 74,17%.

3.4.2 Percobaan Kedua (Data 80:20)

Percobaan kedua dilakukan menggunakan pembagian data 80:20 dengan 198 data training dan 50 data testing. Hasil klasifikasi percobaan kedua dapat dilihat pada tabel 9 berikut ini:

Tabel 9. Uji Coba SVM Percobaan Kedua

<i>Predicted Values</i>	<i>Actual Values</i>	
	True Tepat	True Terlambat
Pred.Tepat	111	69
Pred.Terlambat	4	14

Tabel 9 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). Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan kedua diperoleh nilai *accuracy* sebesar 63,13%, nilai *precision* sebesar 61,67%, nilai *recall* sebesar 96,52% dan nilai *f1-score* sebesar 75,26%.

3.4.3 Percobaan Ketiga (Data 70:30)

Percobaan ketiga dilakukan menggunakan pembagian data 70:30 dengan 174 data training dan 74 data testing. Hasil klasifikasi percobaan ketiga dapat dilihat pada tabel 10 berikut ini:

Tabel 10. Uji Coba SVM Percobaan Ketiga

<i>Predicted Values</i>	<i>Actual Values</i>	
	True Tepat	True Terlambat
Pred.Tepat	92	54
Pred.Terlambat	9	19

Tabel 10 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). 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%.

3.4.4 Percobaan Keempat (Data 60:40)

Percobaan keempat dilakukan menggunakan pembagian data 60:40 dengan 148 data training dan 100 data testing. Hasil klasifikasi percobaan keempat dapat dilihat pada tabel 11 berikut ini:

Tabel 11. Uji Coba SVM Percobaan Keempat

<i>Predicted Values</i>	<i>Actual Values</i>	
	True Tepat	True Terlambat
Pred.Tepat	83	49
Pred.Terlambat	3	13

Tabel 11 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). Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan keempat diperoleh nilai *accuracy* sebesar 64,86%, nilai *precision* sebesar 62,88%, nilai *recall* sebesar 96,51% dan nilai *f1-score* sebesar 76,15%.

3.4.5 Percobaan Kelima (Data 50:50)

Percobaan kelima dilakukan menggunakan pembagian data 50:50 dengan 124 data training dan 124 data testing. Hasil klasifikasi percobaan kelima dapat dilihat pada tabel 12 berikut ini:

Tabel 12. Uji Coba SVM Percobaan Kelima

Predicted Values	Actual Values	
	True Tepat	True Terlambat
Pred.Tepat	68	41
Pred.Terlambat	4	11

Tabel 5.6 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). Berdasarkan hasil pengukuran performa model *Support Vector Machine* pada percobaan kelima diperoleh nilai *accuracy* sebesar 62,39%, nilai *precisión* sebesar 62,39%, nilai *recall* sebesar 94,44% dan nilai *f1-score* sebesar 75,14%.

3.4.6 Performa Support Vector Machine

Setelah melakukan uji coba dengan 5 percobaan maka rata rata performa *Support Vector Machine* berdasarkan nilai *accuracy*, nilai *precisión*, nilai *recall*, dan nilai *f1-score* dapat dilihat pada tabel 13 berikut ini:

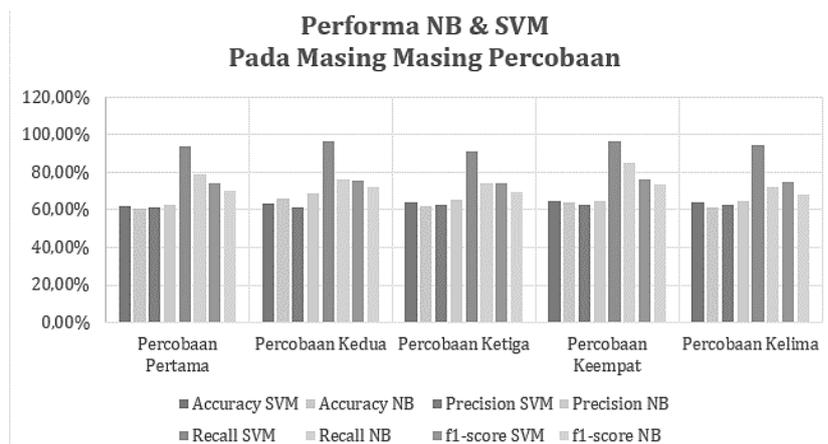
Tabel 13. Performa Support Vector Machine

NO	PERCOBAAN	DATA	PERFORMA SUPPORT VECTOR MACHINE			
			Accuracy	Precision	Recall	f1-score
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		62,88%	65,27%	93,51%	75,25%

Pada tabel 13 diketahui nilai *accuracy* tertinggi diperoleh pada percobaan keempat yaitu 64,86%, nilai *precisión* 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 *accuracy* sebesar 63,51%, *precision* sebesar 62,25%, *recall* sebesar 94,48% dan *f1-score* sebesar 75,04%.

3.3 Pembahasan

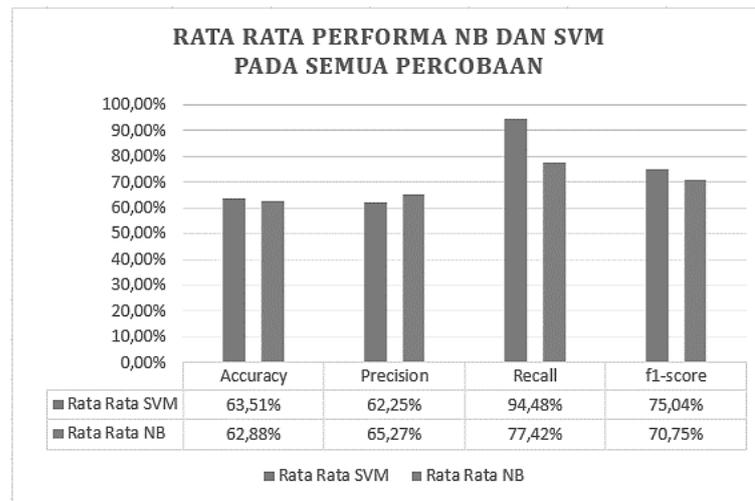
Setelah diketahui hasil pengujian *Naïve Bayes* dan *Support Vector Machine* pada percobaan pertama sampai percobaan kelima, tahap selanjutnya adalah melakukan perbandingan performa *Naïve Bayes* dan *Support Vector Machine* berdasarkan nilai *accuracy*, nilai *precisión*, nilai *recall* dan nilai *f1-score*. Perbandingan performa *Naïve Bayes* dan *Support Vector Machine* dapat dilihat pada gambar 4.



Gambar 4. Perbandingan Performa NB dan SVM

Pada gambar 4 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 5 berikut ini:



Gambar 5. Rata Rata Performa NB dan SVM

Berdasarkan perolehan rata rata performa *Naïve Bayes* dan *Support Vector Machine* pada gambar 5 diketahui bahwa *Support Vector Machine* memperoleh rata rata nilai *accuracy* (63,51%), *precisión* (62,25%), *recall* (94,48%), dan *f1-score* (75,04%). Sedangkan *Naïve Bayes* memperoleh rata rata nilai *accuracy* (62,88%), *precisión* (65,27%), *recall* (77,42%), dan *f1-score* (70,75%).

4. KESIMPULAN

Penelitian ini berhasil melakukan klasifikasi keterlambatan pembayaran SPP menggunakan *Naïve Bayes* dan *Support Vector Machine*. Data yang digunakan adalah data siswa di Madrasah Tsanawiyah Lukmanul Hakim Samarinda tahun ajaran 2021/2022 sampai 2023/2024. Berdasarkan hasil pengukuran performa model *Naïve Bayes* dan *Support Vector Machine* melalui percobaan pertama sampai percobaan kelima diketahui rata rata nilai akurasi *Naïve Bayes* sebesar 62,88% sedangkan *Support Vector Machine* sebesar 63,51%.

REFERENCES

- Abdullah et al. (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.
- Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Energy Universitas Panca Marga*, 8(1).
- 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>
- 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).
- Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika*, 8(1).
- 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>
- 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>
- Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, 24(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565>



- 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).
- Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 6.
- Sukesa, I. K. (2022). *CRISP DM Sebagai Salah Satu Standard untuk Menghasilkan Data Driven Decision Making yang Berkualitas*. Diakses pada 11 Januari 2024 dari <https://www.djkn.kemenkeu.go.id/artikel/baca/15134/CRISP-DM-Sebagai-Salah-Satu-Standard-untuk-Menghasilkan-Data-Driven-Decision-Making-yang-Berkualitas.html>
- 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>