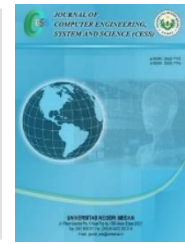


Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Perbandingan Algoritma Naïve Bayes dan Algoritma Support Vector Machine (SVM) Untuk Melihat Potensi Kepatuhan Peserta BPJS Dalam Membayar Tagihan**

***Comparison of Naïve Bayes Algorithm and Support Vector Machine Algorithm to See the Potential of BPJS Participants Compliance in Paying Bills***

Rika Dinda Lestari<sup>1</sup>, Raissa Amanda Putri<sup>2\*</sup>

<sup>1,2</sup> Prodi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara  
Jl.Lap. Golf No.120, Medan Tuntungan, Indonesia  
email: <sup>1</sup>[rkndndstar@gmail.com](mailto:rkndndstar@gmail.com), <sup>2</sup>[raissa.ap@uinsu.ac.id](mailto:raissa.ap@uinsu.ac.id)

**ABSTRAK**

Jaminan Kesehatan Nasional (JKN) yang diselenggarakan oleh Badan Penyelenggara Jaminan Sosial (BPJS) merupakan jaminan sosial yang wajib bagi seluruh warga negara Indonesia berdasarkan Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN). Penelitian ini bertujuan untuk melihat potensi kepatuhan peserta BPJS dalam membayar tagihan dengan menggunakan perbandingan antara algoritma naïve bayes dan algoritma support vector machine (SVM). Pada penelitian ini metode yang digunakan ialah metode kuantitatif dengan menghitung hasil tingkat akurasi dari masing masing algoritma yang digunakan. Sebelum dilakukannya penerapan pada algoritma naïve bayes dan algoritma support vector machine (SVM) maka akan dilakukannya teknik KDD (Knowledge Discovery in Database) agar data yang digunakan lebih akurat. Dengan data sampel sebanyak 1.499 kita dapat mengetahui peserta BPJS mana yang membayar tagihannya sangat tepat waktu, tepat waktu, dan tidak tepat waktu dengan menggunakan teknik klasifikasi dan menerapkan algoritma naïve bayes dan algoritma support vector machine (SVM) pada rapidminer alat sehingga dapat diperoleh hasil akurasi sebesar 99,10%. Dengan nilai sangat tepat waktu sebesar 0,334, tepat waktu dengan nilai sebesar 0,182, dan tidak tepat waktu dengan nilai sebesar 0,484 pada algoritma naïve bayes. Sedangkan pada algoritma support vector machine (SVM) memperoleh tingkat akurasi sebesar 98,48% dengan nilai sangat tepat waktu sebesar 45, tepat waktu dengan nilai 6, dan tidak tepat waktu dengan nilai 45. dilihat dari hasil tingkat akurasi algoritma naïve bayes lebih unggul dibandingkan algoritma support vector machine (SVM).

**Kata Kunci:** BPJS; data mining; klasifikasi; naïve bayes; support vector machine

\*Penulis Korespondensi:  
email: [raissa.ap@uinsu.ac.id](mailto:raissa.ap@uinsu.ac.id)

## ABSTRACT

National Health Insurance (JKN) which is administered by the Social Security Administering Body (BPJS) is mandatory social security for all Indonesian citizens based on Law Number 40 of 2004 concerning the National Social Security System (SJSN). This research aims to see the potential for BPJS participant compliance in paying bills by using a comparison between the Naïve Bayes algorithm and the support vector machine (SVM) algorithm. In this research, the method used is a quantitative method by calculating the results of the accuracy level of each algorithm used. Before implementing the Naive Bayes algorithm and the support vector machine (SVM) algorithm, the KDD (Knowledge Discovery in Database) technique will be carried out so that the data used is more accurate. With 1,499 sample data, we can find out which BPJS participants pay their bills on time, on time and not on time by using classification techniques and applying the Naïve Bayes algorithm and the support vector machine (SVM) algorithm on the Rapidminer tool so that accurate results can be obtained. of 99.10%. With a very timely value of 0.334, timely with a value of 0.182, and not timely with a value of 0.484 in the Naïve Bayes algorithm. Meanwhile, the support vector machine (SVM) algorithm obtained an accuracy level of 98.48% with a very timely value of 45, timely with a value of 6, and not timely with a value of 45. Judging from the results, the accuracy level of the Naïve Bayes algorithm is superior to support vector machine (SVM) algorithm.

**Keywords:** *BPJS; data mining; classification; naive Bayes; support vector machine*

---

## 1. PENDAHULUAN

Jaminan Kesehatan Nasional (JKN) yang diselenggarakan oleh Badan Penyelenggara Jaminan Sosial (BPJS) merupakan jaminan sosial yang wajib bagi seluruh warga negara Indonesia berdasarkan Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN). Setiap anggota masyarakat yang berkontribusi secara mandiri atau didanai oleh pemerintah berhak memperoleh pelayanan kesehatan dasar [1]. Kepesertaan BPJS Kesehatan terbagi menjadi dua, yaitu Peserta Penerima Bantuan Iuran (PBI) dan Peserta Bukan Penerima Bantuan (Non-PBI). Peserta PBI atau peserta yang tidak mampu membayar iuran dibayar oleh pemerintah, sedangkan peserta non-PBI yang mempunyai penghasilan, iuran dibayarkan secara mandiri atau dengan dipotong persentase dari gajinya.

Tanggal pembayaran tagihan BPJS jatuh pada tanggal 10 setiap bulannya. Peserta BPJS yang melakukan pembayaran tagihan sebelum tanggal 10 dapat disebut sebagai peserta BPJS yang tepat waktu dalam menjalankan kewajibannya. Seorang peserta BPJS dikatakan tidak tepat waktu jika membayar tagihannya setelah tanggal 10. Ketidapatuhan dalam membayar tagihan BPJS Kesehatan sering terjadi jika peserta tidak memahami siklus BPJS Kesehatan. Untuk itu kita perlu memiliki pemahaman tentang BPJS Kesehatan agar kita dapat segera mengapresiasinya secara utuh[2]. Dengan bertambahnya jumlah peserta BPJS dan data pembayaran tagihan, BPJS Kesehatan belum membuat pola untuk mengklasifikasikan berapa jumlah peserta BPJS yang sangat tepat waktu, tepat waktu, dan tidak tepat waktu berdasarkan data peserta BPJS. Untuk menemukan pola yang tepat, perlu dilakukan perbandingan antara algoritma naïve Bayes dan algoritma support vector machine (SVM).

Pada penelitian sebelumnya dengan judul "Perbandingan Metode Naive Bayes dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus" penelitian ini menggunakan alat bantu WEKA dengan tools options Cross Validation dan Confussion Matrix dengan hasil akurasi dan klasifikasi dengan 3 class yaitu positif dan negatif. [3] Sedangkan

dalam penelitian ini yang dilakukan saat ini ialah menggunakan alat bantu tools RapidMiner dan mengklasifikasikan 3 class dalam mengklasifikasikan data peserta BPJS dalam membayar tagihan sangat tepat waktu, tepat waktu, dan tidak tepat waktu, diperoleh 1500 data dalam 3 bulan tahun 2023 yaitu Januari, Februari, dan Maret. Setelah dilakukan tahap KDD kini terdapat 1499 data, karena sebelumnya didalamnya terdapat data yang hilang. Untuk menghasilkan tingkat akurasi dari klasifikasi maka diperlukan penelitian perbandingan antara algoritma naïve bayes dengan algoritma support vector machine (SVM) untuk data latih dan 70% data uji untuk 30% data latih. Algoritma yang akan digunakan adalah algoritma naïf bayes dan algoritma svm. Algoritma naïve bayes merupakan algoritma yang banyak digunakan untuk mengklasifikasikan data dan algoritma svm merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi.

## **2. DASAR/TINJAUAN TEORI**

### **2.1 Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan**

Jaminan Kesehatan Nasional (JKN) yang diselenggarakan oleh Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan merupakan jaminan sosial yang wajib bagi seluruh warga negara Indonesia berdasarkan Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN). BPJS dibentuk sesuai dengan amanat Undang-Undang Nomor 24 Tahun 2011 [4]. Sebuah lembaga hukum bernama BPJS Kesehatan didirikan untuk mengawasi inisiatif asuransi kesehatan. Badan hukum bernama Badan Penyelenggara Jaminan Sosial (BPJS) dibentuk untuk mengawasi program jaminan sosial, yang bertujuan untuk memastikan bahwa setiap peserta atau anggota keluarga mempunyai akses terhadap kebutuhan hidup yang layak[1].

### **2.2 Penambangan Data**

Data mining secara sederhana merupakan langkah ekstraksi untuk memperoleh informasi penting yang bersifat implisit dan tidak diketahui[5]. Data mining adalah proses yang menggunakan kecerdasan buatan, teknik statistik, matematika, dan pembelajaran mesin untuk mengidentifikasi dan mengekstrak informasi berguna dan pengetahuan terkait dari berbagai database besar [6]. Selain itu, data mining adalah proses menganalisis informasi yang diperoleh dari pengumpulan data dengan menggunakan berbagai teknik pembelajaran mesin [7]. Teknik, metode, dan algoritma yang digunakan cukup beragam. Pemilihan metode atau algoritma yang efektif mempunyai dampak yang signifikan terhadap tujuan dan proses pengetahuan dan penemuan dalam database (KDD) secara keseluruhan[8]. Metode klasifikasi data mining melibatkan kelompok sasaran secara akurat. Metode klasifikasi ini bertujuan untuk mengelompokkan nilai suatu variabel yang belum diketahui dari variabel lain yang telah ditentukan sebelumnya [9].

### **2.3 Penemuan Pengetahuan dalam Database (KDD)**

Penemuan pengetahuan dalam database (KDD) adalah kata lain yang digunakan dalam data mining yang memiliki arti mirip dengan data mining. Tujuan KDD adalah menggunakan data database dan mengolahnya untuk menghasilkan informasi baru yang berharga. Proses KDD menggunakan data mining untuk mengekstrak pengetahuan tentang apa yang dianggap tepat sesuai spesifikasi ukuran dan batas[10]. Tahapan proses penemuan pengetahuan dalam database (KDD) adalah seleksi data, preprocessing data, transformasi data, penambangan data, dan evaluasi.

## 2.4 Klasifikasi

Klasifikasi adalah teknik yang digunakan untuk menemukan model untuk menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas suatu objek yang labelnya tidak diketahui [5]. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan yaitu membangun model sebagai prototype untuk disimpan sebagai memori dan menggunakan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi terhadap objek data lain sehingga diketahui objek data tersebut termasuk dalam kelas mana. pada model yang telah disimpan[11]. Proses klasifikasi adalah proses menemukan model (fungsi) yang mendeskripsikan dan membedakan kelas atau konsep data dengan tujuan agar dapat digunakan untuk memprediksi kelas objek yang tidak diketahui [12]. Klasifikasi merupakan proses persiapan pencarian model yang mendeskripsikan data dan mengklasifikasikan nya ke dalam kelas-kelas yang ada[13].

## 2.5 Algoritma Naïve Bayes

Algoritma naïve bayes merupakan statistik pengklasifikasi yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas [14]. Naïve bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari kemungkinan klasifikasi terbesar, dengan melihat frekuensi setiap klasifikasi pada data pelatihan[15]. Metode pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas naïve bayes pengklasifikasi didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi mirip dengan pohon keputusan dan jaringan saraf[14]. Algoritma naïve bayes dapat digunakan untuk memperkirakan probabilitas terisinya suatu kelas tertentu [14].

## 2.6 Support vector machine (SVM)

Support vector machine (SVM) adalah metode klasifikasi untuk data linier dan nonlinier. Singkatnya, SVM adalah algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli menjadi dimensi yang lebih tinggi. Teknik ini termasuk dalam metode klasifikasi tipe terbimbing (supervised) karena mempunyai sasaran pembelajaran tertentu. Klasifikasi dilakukan dengan mencari hyperplanes atau garis pemisah (batas keputusan) yang memisahkan suatu kelas dengan kelas lainnya. Dalam konsep ini, SVM mencoba menemukan hyperplane terbaik di antara fungsi yang jumlahnya tidak terbatas. Fungsi tak terbatas dalam pencarian hyperplane dalam metode SVM merupakan suatu keuntungan, dimana pemrosesan akan selalu dapat dilakukan terlepas dari data yang dimilikinya.

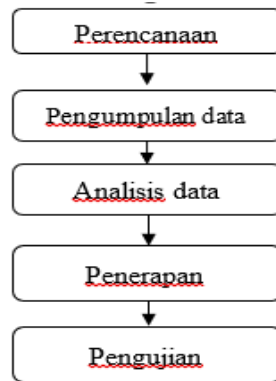
## 2.7 Flow chart

Flowchart adalah gambaran grafis dari langkah-langkah dan urutan prosedural suatu program. Flowchart dapat memberikan gambaran prosedur logika yang efektif, jelas, dan ringkas [16].

## 3. METODE

Penelitian ini menggunakan metode penelitian kuantitatif yang menguji suatu teori dengan menguji hubungan antar variabel. Metode kuantitatif juga bertujuan untuk menguji hipotesis dengan menggunakan teori yang sudah ada. Penelitian yang akan dilakukan adalah membandingkan algoritma naïve Bayes dan algoritma support vector machine (SVM) dengan rapidminer untuk melihat potensi kepatuhan peserta BPJS dalam membayar tagihan. Sebelum dilakukannya penerapan algoritma naïve baye dan algoritma support vector machine (SVM) maka aka dilakukan tahapn KDD terlebih dahulu, agar data yang digunakan lebih akurat dan

tidak terdapat data missing. Berikut langkah-langkah yang akan dilakukan dalam proses penerapan algoritma naïve bayes untuk melihat kepatuhan peserta BPJS dalam membayar tagihan. Kerangka penelitian disajikan pada Gambar 1.



Gambar 1. Kerangka Penelitian

### 3.1 Pengumpulan data

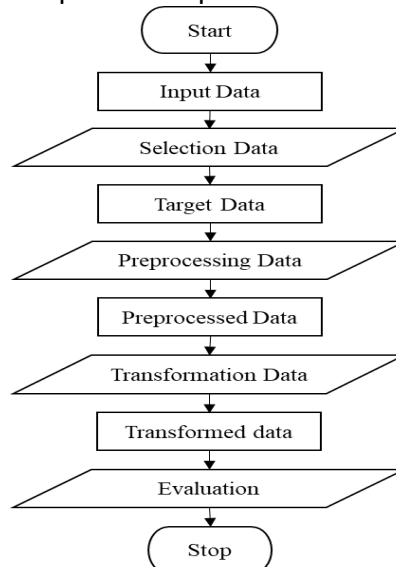
Proses pengumpulan data ini memakan waktu, dimulai dengan menyerahkan surat penelitian dan surat persyaratan kepada instansi, dari bulan Desember lalu hingga bulan April data yang dibutuhkan sudah terkumpul. Peneliti memperoleh 1500 data, 1000 diantaranya akan dijadikan data latih, dan 500 data lagi akan dijadikan data uji. Bentuk datanya dapat dilihat pada Tabel 1.

Tabel 1 Dataset Peserta BPJS Kesehatan

BPJS No	Region	Address	Due date	Payment Date
1719622923	KOTA MEDAN	JL HALAT GG RAMBUNG NO 19 A	02/10/2023	01/02/2023
2738206168	KOTA MEDAN	JL. TITIPAPAN GG. TURI I NO. 10-A	02/10/2023	01/02/2023
7334111	KOTA MEDAN	JL AMALIUN NO 14	02/10/2023	01/02/2023
1150556163	KOTA MEDAN	JL SETIABUDI PSR I GG MAWAR NO 16	02/10/2023	01/02/2023
2478230728	KOTA MEDAN	LINGK07	02/10/2023	01/02/2023

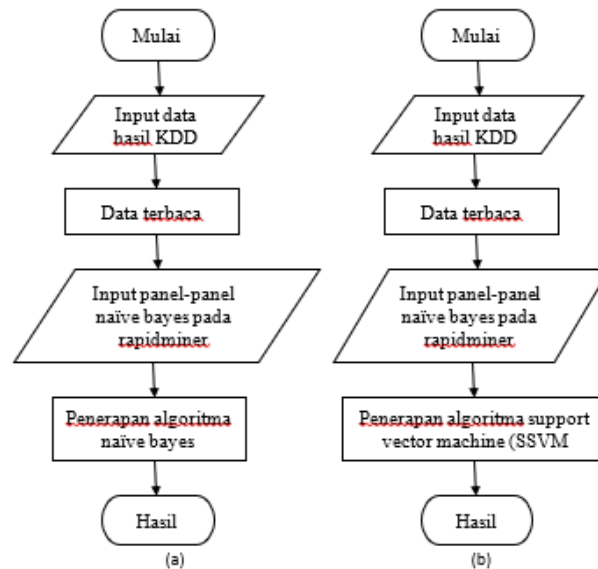
### 3.2. Analisis

Dalam penelitian ini analisis data yang dilakukan adalah pengolahan data dengan menggunakan teknik data mining yaitu Knowledge Discovery in Database (KDD). Alur pemrosesan menggunakan KDD dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Tahapan KDD

Setelah dilakukan teknik data mining maka data yang kita gunakan menjadi lebih akurat dan dapat kita uji pada tools RapidMiner dengan algoritma naïve bayes dan algoritma support vector machine (SVM). Alur pemrosesan menggunakan algoritma naïf Bayes dan algoritma SVM dapat dilihat pada Gambar



**Gambar 3.** Flowchart Implementasi (a) Algoritma Naïve Bayes, (b) Algoritma SVM

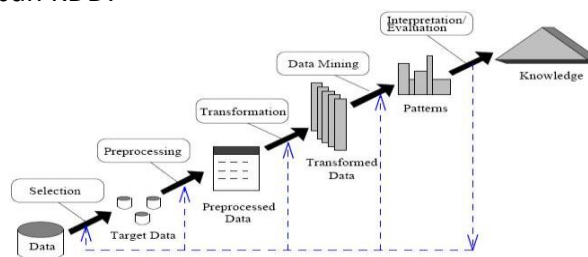
### 3.2 Pengujian

Pengujian pada penelitian ini dilakukan untuk melihat perbandingan nilai akurasi antara algoritma naïve bayes dengan algoritma support vector machine (SVM) untuk melihat potensi kepatuhan peserta BPJS dalam membayar tagihan dengan menggunakan teknik klasifikasi dan menggunakan tools rapidminer dengan data uji dan data latih untuk menghasilkan tingkat akurasi yang tinggi. Data diklasifikasikan menjadi tiga kelas yaitu sangat tepat waktu, tepat waktu, dan tidak tepat waktu.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Pembahasan

Analisis merupakan suatu langkah dalam memahami masalah yang ingin dipecahkan sebelum mengambil tindakan atau keputusan dalam ujian yang akan dibuat. Data yang digunakan dalam penelitian ini adalah data kepesertaan BPJS Kesehatan dalam membayar tagihan yang bersumber dari kantor Deputy BPJS Kesehatan Cabang 1 Kota Medan. Sebelum implementasi dan pengujian, analisis data pada aplikasi data mining menggunakan tahap Knowledge Discover in Database (KDD) yang sering digunakan secara bergantian untuk menggambarkan proses penggalian informasi tersembunyi dalam data berukuran besar. Berikut gambaran tahapan KDD:



**Gambar 4.** Tahapan KDD

#### 4.1.1 Analisis Data

Input merupakan tahap awal untuk menganalisis data. Data yang akan diinput merupakan data mentah yang disediakan oleh BPJS Kesehatan. Pada data mentah yang disediakan, ada beberapa data yang tidak perlu digunakan sehingga perlu diseleksi terlebih dahulu. Sebelum menginputkan data, telah dilakukan split data atau pembagian terlebih dahulu. Terdapat 1000 data uji dan 500 data latih.

#### 4.1.2 Data Selection

Pada tahap pemilihan data, fungsinya untuk memilih data yang sesuai dengan kebutuhan. Dalam penelitian ini data yang cocok untuk mendapatkan hasil yang diinginkan dan sesuai dengan proses adalah nomor BPJS, wilayah, tanggal jatuh tempo, dan tanggal pembayaran serta terdapat tambahan kolom data yaitu perkiraan dan perkiraan (hari) sebagai akurasi. untuk memprediksi potensi kepatuhan.

Row No.	Estimasi	No BPJS	Wilayah	Tgl Jth Tempo	Tgl Bayar	Estimasi ( W...
986	Tidak Tepat ...	2082027543	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
987	Tidak Tepat ...	2081473841	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
988	Tidak Tepat ...	2081107034	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
989	Tidak Tepat ...	2084051272	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
990	Tidak Tepat ...	2085051431	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
991	Tidak Tepat ...	2085033701	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
992	Tidak Tepat ...	2084420777	KOTA BINJAI	Mar 10, 2023	Mar 14, 2023	-4
993	Tepat Waktu	2093557217	KOTA BINJAI	Mar 10, 2023	Mar 10, 2023	0
994	Sangat Tepat...	2199547811	KOTA BINJAI	Mar 10, 2023	Mar 7, 2023	3
995	Sangat Tepat...	2199086076	KOTA BINJAI	Mar 10, 2023	Mar 1, 2023	9
996	Sangat Tepat...	1265066087	KOTA BINJAI	Mar 10, 2023	Mar 5, 2023	5
997	Sangat Tepat...	2200634752	KOTA BINJAI	Mar 10, 2023	Mar 7, 2023	3
998	Sangat Tepat...	2201932124	KOTA BINJAI	Mar 10, 2023	Mar 9, 2023	1
999	Tepat Waktu	2201085415	KOTA BINJAI	Mar 10, 2023	Mar 10, 2023	0
1000	Tidak Tepat ...	2200818993	KOTA BINJAI	Mar 10, 2023	Mar 11, 2023	-1

ExampleSet (1,000 examples, 1 special attribute, 5 regular attributes)

Gambar 5. Hasil Selection Data Uji

Row No.	Estimasi	No BPJS	Wilayah	Tgl Jth Tempo	Tgl Bayar	Estimasi ( W...
486	Tidak Tepat ...	7313714	KOTA MEDAN	Feb 10, 2023	Feb 19, 2023	-9
487	Tidak Tepat ...	7314243	KOTA MEDAN	Feb 10, 2023	Feb 19, 2023	-9
488	Sangat Tepat...	7316853	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
489	Sangat Tepat...	7318517	KOTA MEDAN	Feb 10, 2023	Feb 2, 2023	8
490	Sangat Tepat...	7321465	KOTA MEDAN	Feb 10, 2023	Feb 3, 2023	7
491	Sangat Tepat...	7321904	KOTA MEDAN	Feb 10, 2023	Feb 4, 2023	6
492	Sangat Tepat...	7322264	KOTA MEDAN	Feb 10, 2023	Feb 5, 2023	5
493	Sangat Tepat...	7322657	KOTA MEDAN	Feb 10, 2023	Feb 6, 2023	4
494	Sangat Tepat...	7324413	KOTA MEDAN	Feb 10, 2023	Feb 7, 2023	3
495	Sangat Tepat...	1872268942	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
496	Sangat Tepat...	7325381	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
497	Sangat Tepat...	7326426	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
498	Sangat Tepat...	7328812	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
499	Sangat Tepat...	7333604	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
500	Sangat Tepat...	764256	?	Feb 11, 2023	Feb 2, 2023	9

ExampleSet (500 examples, 1 special attribute, 5 regular attributes)

Gambar 6. Hasil Selection Data Latih

### 4.1.3 Data Preprocessing

Data yang telah diseleksi pada tahap sebelumnya masih terdapat data yang kosong, ganda, dan hilang. Dengan mengolah data yang telah dipilih sebelumnya maka akan terlihat apakah data tersebut mengandung data kosong dan nilai yang hilang. Jika tidak ada data yang kosong dan data yang hilang maka akan segera diproses pada tahap transformasi data, sedangkan jika ada data yang kosong dan data yang hilang akan dibersihkan terlebih dahulu dengan menggunakannya sebelum masuk pada tahap transformasi data.

Label		Least	Most	Values
Estimasi	Nominal	0	Sangat Tepat Waktu (1... Tidak Tepat Waktu (242)	Tidak Tepat Waktu (242), Tepat Waktu (133), ...[1
No BPJS	Real	0	Min: 764256	Max: 3047540736 Average: 293891153.862
Wilayah	Nominal	1	Least: KOTA MEDAN (499)	Most: KOTA MEDAN (499) Values: KOTA MEDAN (499)
Tgl Jth Tempo	Date-time	0	Earliest date: Jan 10, 2023	Latest date: Feb 11, 2023 Duration: 32 days
Tgl Bayar	Date-time	0	Earliest date: Jan 1, 2023	Latest date: Feb 25, 2023 Duration: 55 days
Estimasi ( Waktu )	Integer	0	Min: -21	Max: 9 Average: -2.766

Gambar 7. Data Missing Pada Data Latih

Terlihat pada gambar diatas terdapat satu data yang hilang pada data latih, sehingga perlu dihilangkan dari data tersebut agar lebih akurat.

Row No.	Estimasi	No BPJS	Wilayah	Tgl Jth Tempo	Tgl Bayar	Estimasi ( W...
485	Tidak Tepat ...	7313409	KOTA MEDAN	Feb 10, 2023	Feb 19, 2023	-9
486	Tidak Tepat ...	7313714	KOTA MEDAN	Feb 10, 2023	Feb 19, 2023	-9
487	Tidak Tepat ...	7314243	KOTA MEDAN	Feb 10, 2023	Feb 19, 2023	-9
488	Sangat Tepat...	7316853	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
489	Sangat Tepat...	7318517	KOTA MEDAN	Feb 10, 2023	Feb 2, 2023	8
490	Sangat Tepat...	7321465	KOTA MEDAN	Feb 10, 2023	Feb 3, 2023	7
491	Sangat Tepat...	7321904	KOTA MEDAN	Feb 10, 2023	Feb 4, 2023	6
492	Sangat Tepat...	7322264	KOTA MEDAN	Feb 10, 2023	Feb 5, 2023	5
493	Sangat Tepat...	7322657	KOTA MEDAN	Feb 10, 2023	Feb 6, 2023	4
494	Sangat Tepat...	7324413	KOTA MEDAN	Feb 10, 2023	Feb 7, 2023	3
495	Sangat Tepat...	1872268942	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
496	Sangat Tepat...	7325381	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
497	Sangat Tepat...	7326426	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
498	Sangat Tepat...	7328812	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9
499	Sangat Tepat...	7333604	KOTA MEDAN	Feb 10, 2023	Feb 1, 2023	9

ExampleSet (499 examples, 1 special attribute, 5 regular attributes)

Gambar 8. Hasil Preprocessing Data

Gambar diatas adalah hasil pembersihan data latih yang kosong dan hilang pada data BPJS Kesehatan, sudah tidak ada lagi data yang hilang dan kosong, sehingga data BPJS Kesehatan berjumlah 499 data.

### 4.1.4 Data Transformation

Setelah membersihkan data yang hilang, pada tahap ini kita akan melakukan normalisasi data menggunakan panel normalize pada RapidMiner agar data yang digunakan lebih cocok untuk diaplikasikan dengan algoritma naïve bayes.



### Z-Transformation

```

Normalize 2 attributes to mean 0 and variance 1.
Using
No BPJS --> mean: 7.40934261038E8, variance: 1.00094961623943424E18
Estimasi ( Waktu ) --> mean: -1.623, variance: 42.905776776776776
    
```

### Z-Transformation

```

Normalize 2 attributes to mean 0 and variance 1.
Using
No BPJS --> mean: 2.93891153862E8, variance: 5.2911777622010426E17
Estimasi ( Waktu ) --> mean: -2.766, variance: 55.97920240480962
    
```

**Gambar 9.** Hasil Transformasi pada Data Uji dan Data Latih

Pada gambar di atas terlihat data yang telah diubah menjadi bentuk yang sesuai untuk data mining pada tahap selanjutnya.

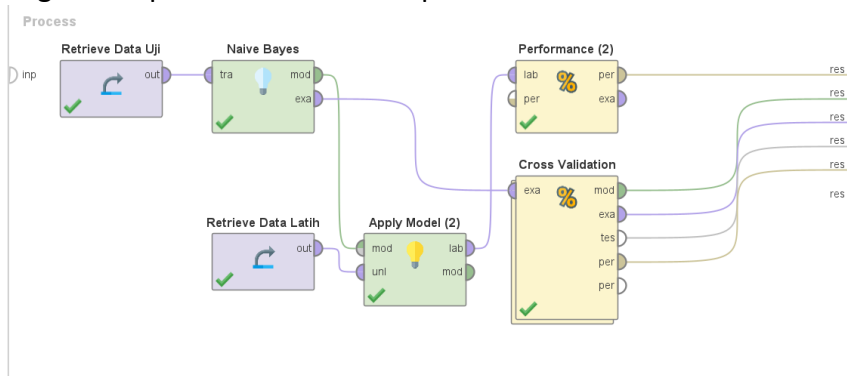
#### 4.1.5 Evaluation

Pada tahap ini dihasilkan pola karakteristik dari model klasifikasi yang dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai.

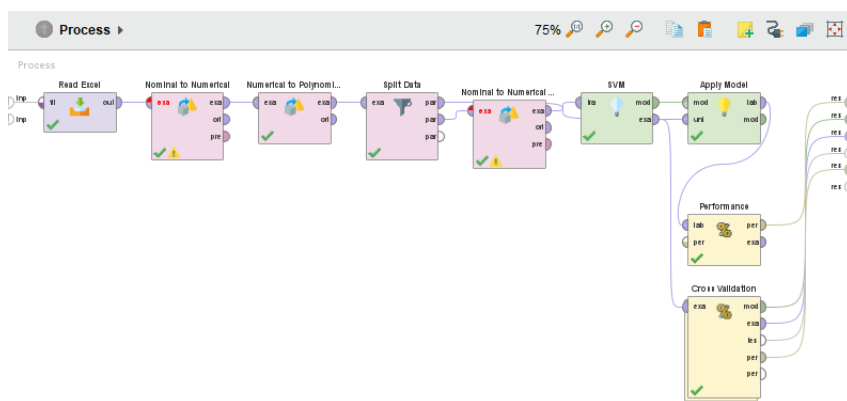
### 4.2 Hasil

#### 4.2.1 Penerapan Algoritma

Penerapan algoritma naïve bayes menggunakan tools RapidMiner untuk menentukan potensi dari kepatuhan peserta BPJS Kesehatan dalam membayar tagihan. Gambar berikut merupakan hasil dari potensi kepatuhan peserta BPJS Kesehatan dengan ditambahkan nilai confidence sebagai hasil prediksi dari tools RapidMiner.



**Gambar 5.** Model Penerapan Naive Bayes dengan RapidMiner



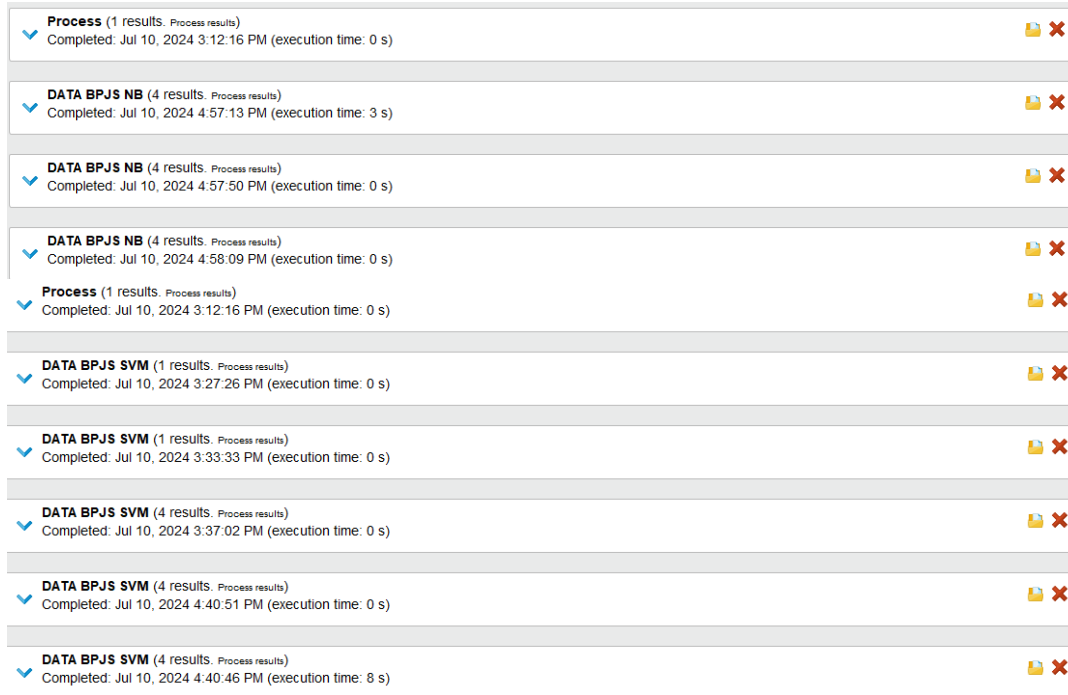
**Gambar 6.** Model penerapan algoritma support vector machine (SVM) dengan RapidMiner

## 4.2.2 Pengujian

Pada tahap ini pola-pola hasil akhir dari proses sebelumnya akan dipresentasikan dalam bentuk visual untuk membantu *user* dalam mengerti pola tersebut.

### 4.2.2.1 Akurasi Prediksi

Pada percobaan dengan algoritma naïve bayes dengan menggunakan tools RapidMiner diperoleh waktu eksekusi sebesar 0 second dan 3 second pada algoritma naïve bayes dapat diartikan eksekusi berjalan cukup cepat. Dan diperoleh waktu 0 second dan 8 second untuk algoritma support vector machine (SVM). Hasil eksekusi dapat dilihat pada gambar berikut.



**Gambar 7.** Hasil eksekusi Naive Bayes dan SVM

Hasil dari tingkat akurasi algoritma naïve bayes sebesar 99.10% yang mengartikan algoritma naïve bayes memiliki tingkat kekuatan yang sangat tinggi. Hal ini dibuktikan dengan hasil akurasi perhitungan mencapai 99.10%, nilai yang tinggi tersebut juga disebabkan oleh akurat nya data uji dan data latih yang mengakibatkan pola tersebut dapat memprediksi dengan akurat dan tepat. Hasil dari Tingkat akurasi algoritma support vector machine (SVM) Sebesar 98.48% yang mengartikan algoritma support vector machine (SVM) cukup baik dan tinggi, namun jika dibandingkan dengan algoritma naïve bayes, maka algoritma naïve bayes lebih unggul dari pada algoritma support vector machine (SVM). Adapun hasil akurasi dapat dilihat pada gambar berikut.

accuracy: 99.10% +/- 1.29% (micro average: 99.10%)

	true Sangat Tepat Waktu	true Tidak Tepat Waktu	true Tepat Waktu	class precision
pred. Sangat Tepat Waktu	325	0	0	100.00%
pred. Tidak Tepat Waktu	9	484	0	98.17%
pred. Tepat Waktu	0	0	182	100.00%
class recall	97.31%	100.00%	100.00%	

**Gambar 8.** Hasil Akurasi Naive Bayes

accuracy: 98.48%

	true Tidak Tepat Waktu	true Sangat Tepat Waktu	true Tepat Waktu	class precision
pred. Tidak Tepat Waktu	508	16	0	96.95%
pred. Sangat Tepat Waktu	0	305	0	100.00%
pred. Tepat Waktu	0	0	221	100.00%
class recall	100.00%	95.02%	100.00%	

**Gambar 9.** Hasil akurasi SVM

```

SimpleDistribution

Distribution model for label attribute Estimasi

Class Sangat Tepat Waktu (0.334)
5 distributions

Class Tidak Tepat Waktu (0.484)
5 distributions

Class Tepat Waktu (0.182)
5 distributions
    
```

**Gambar 10.** Simple distributions naive bayes

Dari gambar diatas dapat ditarik kesimpulan bahwa data BPJS Kesehatan Kedeputusan Wilayah 1 Cabang Kota Medan, kelas tidak tepat waktu dengan nilai sebesar 0.484 lebih besar dari kelas sangat tepat waktu dengan nilai 0.334 dan kelas tepat waktu dengan nilai 0.182. Adapun model distribusi untuk label atribut kelas estimasi adalah sebagai berikut:

- Kelas sangat tepat waktu : 5 distributions
- Kelas tepat waktu : 5 distributions
- Kelas tidak tepat waktu : 5 distributions

### Kernel Model

```

Total number of Support Vectors: 96
Bias (offset): -0.875

Feature weight calculation only possible for two class learning problems.
Please use the operator SVMWeighting instead.

number of classes: 3
number of support vectors for class Tidak Tepat Waktu: 45
number of support vectors for class Tepat Waktu: 6
number of support vectors for class Sangat Tepat Waktu: 45
    
```

**Gambar 11.** Kernel model SVM

Dari gambar diatas dapat ditarik kesimpulan bahwa data BPJS Kesehatan Kedeputusan Wilayah 1 Cabang Kota Medan, kelas tidak tepat waktu dengan nilai sebesar 45 sama dengan kelas sangat tepat waktu dengan nilai 45 dan kelas tepat waktu dengan nilai 6.

#### 4.2.2.2 Percobaan Perhitungan Manual

Untuk memastikan bahwa penerapan algoritma diatas benar akurat, akan dilakukan perhitungan manual untuk salah satu algoritma yaitu algoritma Naïve Bayes. Perhitungan

manual yang dilakukan dengan data uji sebanyak 1000 data. Berikut perhitungan manual dengan algoritma naïve bayes:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Keterangan:

$P(H|E)$  = Probabilitas

$H$  = Hipotesis

$E$  = Bukti

P1 = Sangat tepat waktu

P2 = Tepat waktu

P3 = Tidak tepat waktu

Menghitung jumlah kelas dari estimasi berdasarkan klasifikasi yang terbentuk (*prior probability*), dimana jumlah tiap kelas masing-masing estimasi yang ada dibagi dengan jumlah seluruh data.

- a. P1 Class estimasi = "sangat tepat waktu" = jumlah P1  
= 334/1000  
= 0.334
- b. P2 Class Estimasi = "tepat waktu" = jumlah P2  
= 182/1000  
= 0.182
- c. P3 Class Estimasi = "tidak tepat waktu" = jumlah P3  
= 484/1000  
= 0.484

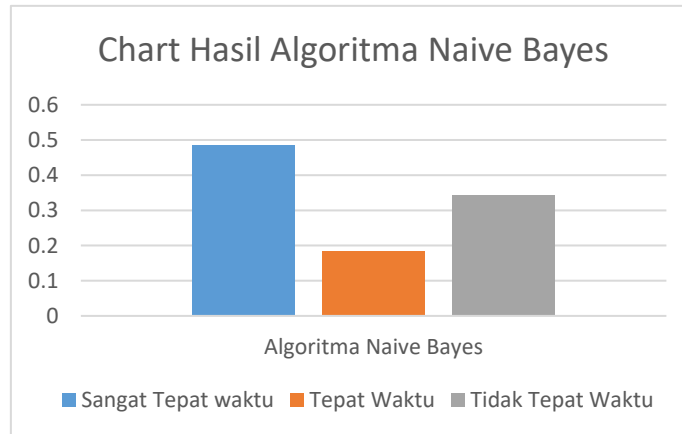
#### 4.2.2.3 Hasil Pengujian

Setelah melakukan penelitian telah mendapat hasil dari penelitian tersebut berupa informasi dan pengetahuan baru dalam menentukan kepatuhan peserta BPJS dalam membayar tagihan sangat tepat waktu, tepat waktu dan tidak tepat waktu. Setelah dilakukan tahapan KDD (Knowledge Discovery in Database) dari data peserta BPJS Kesehatan Kedeputusan 1 cabang kota medan tahun 2023. Dari penelitian tersebut didapatkan hasil yang sesuai dengan tujuan untuk menghasilkan informasi dan mengetahui perbandingan antara algoritma naïve bayes dan algoritma support vector machine (SVM) tentang kepatuhan peserta BPJS Kesehatan yang berpotensi sangat tepat waktu, tepat waktu dan tidak tepat waktu.

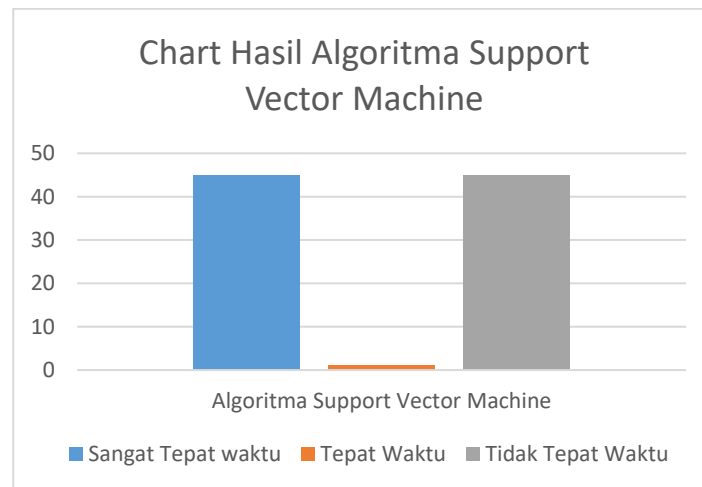
Dari proses pengujian menggunakan tools RapidMiner dengan Teknik klasifikasi didapatkan hasil perbandingan antara algoritma naïve bayes dan algoritma support vector machine (SVM) dengan nilai akurasi 99.10% untuk naïve bayes dan 98.48% untuk SVM. Dari nilai akurasi tersebut dapat disimpulkan bahwa algoritma naïve bayes lebih unggul daripada algoritma support vector machine (SVM). Untuk nilai dari masing-masing kelas diperoleh nilai 0.484 untuk sangat tepat waktu, 0.182 untuk tepat waktu dan 0.343 untuk tidak tepat waktu. Dapat dilihat bahwa peserta BPJS Kesehatan berpotensi tidak tepat waktu dalam membayar tagihan, pada algoritma naïve bayes. Untuk algoritma support vector machine (SVM) diperoleh nilai 45 untuk sangat tepat waktu, 6 untuk tepat waktu dan 45 untuk tidak tepat waktu, disini peserta BPJS Kesehatan berpotensi sangat tepat waktu dan tidak tepat waktu. salah satu hal yang mungkin membuat peserta BPJS membayar dengan tidak tepat waktu ialah besarnya tagihan BPJS Kesehatan tersebut dan juga faktor ekonomi yang mengakibatkan peserta BPJS membayar dengan tidak tepat waktu.

**Tabel 2.** Hasil Pengujian

Algoritma	Nilai Akurasi	Estimasi	Nilai
Naïve bayes	99.10%	Sangat tepat waktu	0.484
		Tepat waktu	0.182
		Tidak tepat waktu	0.343
Support vector machine (SVM)	98.48%	Sangat tepat waktu	45
		Tepat waktu	6
		Tidak tepat waktu	45



**Gambar 12.** Chart Algoritma Naive Bayes



**Gambar 18.** Chart Algoritma Support Vector Machine

## 5. KESIMPULAN

Perbandingan antara Algoritma Naïve Bayes dan Algoritma Support Vector Machine (SVM) untuk melihat potensi kepatuhan peserta BPJS dalam membayar tagihan telah berhasil dilakukan dengan bantuan dari tools RapidMiner. Berdasarkan pengujian tersebut algoritma naïve bayes lebih unggul dari pada algoritma support vector machine (SVM). Dari hasil observasi terhadap data BPJS Kesehatan Kedeputusan wilayah 1 cabang medan melalui teknik klasifikasi menggunakan algoritma naïve bayes didapatkan hasil yang berpotensi tidak tepat waktu terdapat pada sebagian peserta BPJS Kesehatan tersebut dengan nilai sebanyak 0.484 dengan Tingkat akurasi sebesar 99.10%. hal ini disebabkan oleh besarnya jumlah total tagihan

yang menunggak yang membuat semakin tingginya jumlah tagihan menyebabkan sulitnya masyarakat membayar iuran BPJS Kesehatan perbulannya. Sedangkan klasifikasi dengan algoritma support vector machine (SVM) didapat hasil berpotensi sangat tepat waktu dan tidak tepat waktu dengan nilai sebanyak 45 dan Tingkat akurasi sebesar 98.48%

### UCAPAN TERIMA KASIH

Penulis ingin mengungkapkan terimakasih kepada kedua orang tua tercinta Bapak Siswandi dan Ibu Ponimi yang telah memberikan cinta dan kasih sayang dengan membesarkan, mendidik, memberikan semangat dan dukungan serta doa tulus dan perhatian moril maupun materil yang tidak putus. Dalam kesempatan ini penulis juga banyak berterima kasih atas bantuan dan bimbingan yang sangat bermanfaat dari segala pihak. Untuk itu penulis mengucapkan terima kasih sebesar – besarnya kepada Dosen Pembimbing saya Raissa Amanda Putri, S. Kom, M.TI. Tak lupa penulis juga berterima kasih kepada Abang dan Adik serta Keluarga Besar dikampung dan Teman-teman seperjuangan saya yang telah menemani perjalanan selama menempuh pendidikan di Universitas Islam Negeri Sumatera Utara.

### REFERENSI

- [1] M. S. Haq, S. A. Fachrin, and M. K. Alwi, "Faktor Yang Berhubungan Dengan Kepatuhan Membayar Iuran BPJS Peserta Mandiri Di Puskesmas Rappokalling Kota Makassar," *Wind. Public Heal. J.*, vol. 3, no. 4, pp. 615–623, 2022, doi: 10.33096/woph.v3i4.464.
- [2] F. Fajrini, N. Latifah, D. Hermansyah, and N. Firda, "Studi Ketidapatuhan Membayar Iuran BPJS Bukan Penerima Upah di Kelurahan Cempaka Putih," *Public Heal. J.*, 2021, [Online]. Available: <https://media.neliti.com/media/publications/227635-peran-umkm-dalam-pembangunan-dan-kesejah-7d176a2c.pdf%0Ahttp://journal.uinsgd.ac.id/index.php/jispo/article/view/1752>
- [3] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year2020.page133-143.
- [4] M. Novita, "Faktor-Faktor Yang Berhubungan Dengan Kepatuhan Membayar Iuran BPJS Kesehatan Mandiri pada Pasien Rawat Jalan di Puskesmas Lubuk Basung Kab. Agam," *Study Fenomenol. Pengalaman Kel. Pasien Dalampenanganan Prehospital Pasca Deteksi Dini Stroke*, vol. 11, no. 2, pp. 282–289, 2020.
- [5] M. A. M. B. Prasetyo, E. L. H. M. A. J. H. Mirqotussa'adah, and S. H. R. A. Nurzahputra, *Data Mining Algoritma C4.5*. 2019.
- [6] R. A. Kurniawan, M. S. Hasibuan, P. Piramida, and R. S. Ramadhan, "Penerapan Algoritma K-Means Untuk Clustering Tempat Makan Di Batubara," *J. Comput. Sci. Informatics Eng.*, vol. 01, no. 1, pp. 10–18, 2022, doi: 10.55537/cosie.v1i1.27.
- [7] A. Rahman, "Klasifikasi Performa Akademik Siswa Menggunakan Metode Decision Tree dan Naive Bayes," *J. SAINTEKOM*, vol. 13, no. 1, pp. 22–31, 2023, doi: 10.33020/saintekom.v13i1.349.
- [8] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [9] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes," *J. SAINTEKOM*, vol. 13, no. 1, pp. 42–54, 2023,

doi: 10.33020/saintekom.v13i1.352.

- [10] Q. Widayati, "Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk Melihat Potensi Kepatuhan Wajib Pajak Bumi Dan Bangunan," *J. Ilm. Matrik*, vol. 20, no. 2, pp. 157–168, 2019, doi: 10.33557/jurnalmatrik.v20i2.119.
- [11] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [12] F. D. Pratama, I. Zufria, and T. Triase, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 77–84, 2022, doi: 10.36341/rabit.v7i1.2217.
- [13] E. Undamayanti *et al.*, "Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 2, pp. 916–930, 2022.
- [14] M. Furqan, Y. R. Nasution, and R. Fadillah, "Klasifikasi Penyakit Kulit Menggunakan Algoritma Naïve Bayes Berdasarkan Tekstur Warna Berbasis Android," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 12–20, 2022.
- [15] H. Hermanto, A. Mustopa, and A. Y. Kuntoro, "Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 211–220, 2020, doi: 10.33480/jitk.v5i2.1181.
- [16] I. A. Dianta, *Logika dan Algoritma*. Semarang: Yayasan Prima Agus Teknik Redaksi, 2021.