

## **PERBANDINGAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE* DALAM KLASIFIKASI PENYAKIT DIABETES MELITUS (STUDI KASUS : RS. SITI KHADIJAH PALEMBANG)**

**Hilda Apriyani<sup>1</sup>, Kurniati<sup>2</sup>**

Fakultas Ilmu Komputer, Universitas Bina Darma

Email: [aprianihilda9@gmail.com](mailto:aprianihilda9@gmail.com)<sup>1)</sup>, [Kurniati@binadarma.ac.id](mailto:Kurniati@binadarma.ac.id)<sup>2</sup>

### **ABSTRAK**

Diabetes melitus merupakan penyakit kronis yang terjadi akibat kadar glukosa didalam darah yang terlalu tinggi sehingga tidak adanya insulin. Dalam kurun waktu data di Rumah Sakit Islam Siti Khadijah Palembang yang dipengaruhi oleh jumlah dari pasien yang melakukan pemeriksaan kesehatan seperti penyakit diabetes melitus sehingga berpengaruh dalam hal klasifikasi data yang akan menyulitkan pihak rumah sakit. Maka dengan memanfaatkan data mining, pengklasifikasian untuk menentukan pasien yang telah melakukan pemeriksaan termasuk penderita penyakit diabetes atau tidak. Dengan adanya permasalahan tersebut maka penulis melakukan analisis perbandingan dari dua algoritma yaitu algoritma *naïve bayes* dan algoritma *support vector machine* untuk klasifikasi penyakit diabetes dengan menggunakan alat bantu WEKA dengan tools options Cross Validation dan Confusion Matrix dengan hasil akurasi tertinggi yaitu algoritma *support vector machine* dengan kernel *polynomial* yang hasilnya 96.2704% dan tingkat error sebanyak 3.7296% dapat disimpulkan algoritma yang akurat dalam klasifikasi penyakit diabetes yaitu algoritma *support vector machine* dengan kernel *polynomial*.

**Kata Kunci:** Data Mining, Diabetes Melitus, *Naïve Bayes*, *Support Vector Machine*, WEKA.

### **1. PENDAHULUAN**

Diabetes melitus atau di Indonesia lebih dikenal dengan istilah kencing manis. Diabetes melitus merupakan penyakit yang mengakibatkan terlalu banyak kadar gula didalam darah tidak terkendali sehingga tidak adanya insulin oleh sebab itu tubuh tidak berjalan secara normal.

Rumah sakit Siti Khadijah Palembang merupakan salah satu rumah sakit swasta yang telah terakreditasi tipe B rumah sakit ini beralamatkan di jalan Demang Lebar Daun Lorok Pakjo, Kec. Ilir Barat I Kota Palembang Sumatera Selatan. Dalam kurun waktu dengan banyaknya pasien yang melakukan pemeriksaan kesehatan sehingga berdampak pada penumpukkan jumlah data dirumah sakit untuk hasil pemeriksaan tersebut dapat mendiagnosa pasien berpotensi menderita penyakit diabetes melitus sehingga menyulitkan pihak rumah sakit dalam mengklasifikasikan data. Akan hal itu, sangat diperlukan teknik yang dapat mendukung dalam menyelesaikan masalah di atas dengan menerapkan data *mining* untuk klasifikasi dengan memanfaatkan algoritma yaitu Algoritma *Naive Bayes* dan *Support Vector Machine*.

Sebelumnya ada sejumlah penelitian terdahulu yang dilakukan oleh beberapa peneliti seperti perbandingan algoritma *Naive Bayes* dan *Support Vector Machine* diantaranya adalah penelitian yang dilakukan (Riyanto, 2018) untuk pengklasifikasikan jumlah pembaca *online* dengan perbandingan metode *Naive Bayes* dan *SVM*, yang menghasilkan tingkat akurasi yang tepat yaitu algoritma *SVM* dengan nilai **63,39%**. Selanjutnya penelitian yang dilakukan oleh (Arifin & Sasongko, 2018) membandingkan algoritma *Naive Bayes* dan *SVM* untuk klasifikasi jalur minat SMA yang hasil akurasinya, metode *SVM* lebih akurat dibanding algoritma *Naive Bayes* dengan nilai akurasi **97.01%**. Ada lagi penelitian yang dilakukan (Widyawati & Sutanto, 2019) yaitu membandingkan algoritma *Naive Bayes* dan *SVM* dalam klasifikasi SMS *spam* berbahasa Indonesia yang hasil akurasinya, algoritma *Naive Bayes* lah yang lebih unggul dalam hal recall

94% dan Presisi 95%. Sehingga akan muncul pertanyaan pada penelitian ini “Algoritma mana yang tingkat akurasi lebih tepat dalam mengklasifikasi penyakit diabetes melitus antara *Naive Bayes* dan *Support Vector Machine*?”.

Pada penelitian ini data yang digunakan data tahun 2017-2019 sebanyak 613 record. Tujuan dari penelitian ini untuk mengetahui algoritma mana yang tingkat akurasi yang paling akurat untuk klasifikasi penyakit diabetes melitus antara algoritma *Naive Bayes* dan *Support Vector Machine* adalah tujuan dari penelitian ini. sertamanfaat dari penelitian ini yaitu peneliti dapat memahami konsep data *mining* serta bisa bermanfaat bagi peneliti selanjutnya sebagai *referensi* penelitian-penelitian berikutnya yang memiliki topik pembahasan yang sama.

## 2. TINJAUAN PUSTAKA

### 2.1 Data Mining

Menurut penelitian dari(Saleh & Utama, 2016) data *mining* merupakan suatu proses ekstraksi *knowledge* secara otomatis atau penganalisisan data untuk mencari informasi yang mempekerjakan satu atau lebih teknik *machine learning*. Secara garis besar data *mining* yaitu sebuah proses atau metode untuk pencarian data secara otomatis untuk memperoleh informasi yang berfungsi untuk tempat penyimpanan data yang berukuran besar (*bigdata*). Metode untuk tahap mengklasifikasi penyakit diabetes melitus digunakan metode klasifikasi dengan tahap ciri dan atribut dari data yang telah didefinisikan. Algoritma yang akan digunakan dalam proses klasifikasi yaitu algoritma *naive bayes* dan *support vector machine*.

### 2.2 Diabetes Melitus

Menurut penelitian dari(Nugroho et al., 2017) Diabetes Mellitus (DM) atau di Indonesia lebih dikenal dengan istilah kencing manis adalah suatu gangguan atau penyakit yang disebabkan oleh kenaikan gula darah akibat penurunan sekresi insulin atau ketidakseimbangan tubuh dalam memanfaatkan insulin, sehingga kadar glukosa atau kadar gula dalam darah tidak terkendali. Diabetes tipe 2 merupakan penyakit hiperglikemi atau gangguan metabolic yang ditandai dengan naiknya kadar gula darah akibat penurunan sekresi insulin.

### 2.3 Algoritma Naive Bayes

Dalam penelitian (Andriyanto & Santoso, 2017) Algoritma *Naive Bayes* merupakan metode yang memanfaatkan probabilitas dan statistik untuk menyelesaikan masalah klasifikasi. Metode ini melakukan klasifikasi dengan menghitung nilai dari probabilitas  $P(x|y)$  dengan mengetahui probabilitas kelas X. penentuan untuk kelas dalam klasifikasi dilakukan dengan memilih nilai max dari  $P(x|y)$  berdasarkan probabilitas. Keuntungan dari klasifikasi adalah membutuhkan sejumlah data pelatihan dengan relatif kecil untuk memperkirakan parameter yang diperlukan untuk klasifikasi. Berdasarkan algoritma *Naive Bayes* berikut ini persamaan untuk menghitung nilai  $P(x|y)$ :

$$P(X | Y) = \frac{P(Y|X).(X)}{P(Y)} \quad (1)$$

dimana :

$P(X|Y)$  = Posterior|probability yaitu nilai probabilitas X berdasarkan kondisi Y

$P(Y|X)$  = probabilitas Y yang ditentukan X adalah benar

$P(X)$  = Peluang *evidence* penyakit X

$P(Y)$  = Probabilitas dari nilai Y

### 2.4 Algoritma Support Vector Machine

Menurut penelitian(Sukmawati et al., 2013)*Support Vector Machine* (SVM) adalah suatu metode atau algoritma untuk melakukan klasifikasi maupun prediksi. Prinsip kerja dari metode ini adalah mencari ruang pemisah yang paling optimal dari suatu *dataset* dalam kelas yang berbeda. Dalam kehidupan sehari-hari, kita sering diperhadapkan pada persoalan-persoalan yang tidak

*linear* atau data yang tidak dapat benar-benar dipisahkan secara *linear* yaitu suatu kondisi dimana tidak ada sebuah garis atau bidang yang dapat dibuat untuk menjadi pemisah antar kelas data.

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i \vec{x}_i \vec{x}_d + b \tag{2}$$

Dimana:

- $ns$  = Jumlah *support vector*
- $\alpha_i$  = Nilai bobot setiap titik data
- $y_i$  = Kelas data
- $\vec{x}_i$  = Variabel *support vector*
- $\vec{x}_d$  = Data yang akan diklasifikasikan
- $b$  = Nilai error atau bias

### 2.5 K-Fold Cross Validation

Menurut penelitian (Arifin & Sasongko, 2018) bahwa *K-folds cross validation* adalah suatu teknik validasi untuk pengujian yang berfungsi sebagai untuk menilai kinerja proses sebuah algoritma serta digunakan untuk mengurangi waktu dalam proses klasifikasi dengan tetap menjaga nilai keakuratan dengan membagi data secara acak dan mengelompokkan data sebanyak *K-Folds*. Fungsi dari *K-folds cross validation* yaitu agar tidak ada overlapping pada data testing. Berikut untuk ilustrasi sederhana dari *K-folds cross validation* dapat dilihat pada gambar 1 dibawah ini:

|   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Gambar 1. Ilustrasi K-Fold Cross Validation[1]

### 2.6 Confusion matrix

(Rahman et al., 2017) berpendapat bahwa *Confusion matrix* merupakan suatu metode untuk memberikan informasi hasil dari klasifikasi yang dilakukan oleh sistem yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. Sebagai contoh untuk *confusion matrix* dua kelas, maka akan di sebutkan menjadi kelas positif dan kelas negatif. *True positif* mengarah pada kelas positif yang diberi tanda dengan tepat oleh *classifier*, sementara *true negatif* adalah kelas negatif yang diberi tanda dengan tepat oleh *classifier*. Untuk *false positif* adalah kelas negatif yang diberi tanda dengan tidak tepat. selanjutnya, *false negatif* adalah kelas positif yang diberi label dengan tidak tepat.

Tabel 1. Confusion matrix

| Klasifikasi benar | Klasifikasi        |                    |
|-------------------|--------------------|--------------------|
|                   | Positif (+)        | Negatif (-)        |
| Positif benar     | True Positif (TP)  | False Negatif (FN) |
| Negatif benar     | False Positif (FP) | True Negatif (TN)  |

Keterangan :

- True Positif (TP)*, Total daridatapositif yang diklasifikasikan sebagai nilai positif oleh sistem.
- False Positif (FP)*, Total daridata negatif yang diklasifikasikansebagai nilai positif oleh sistem.
- False Negatif (FN)*, Total dari datapositif yang diklasifikasikan nilai positif oleh sistem.
- True Negatif (TN)*, Total dari data negatif yang diklasifikasikannilai negatif oleh sistem.

$$\text{Akurasi} = \frac{(TP + TN)}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$\text{Pesisi} = \frac{(TP + TN)}{TP+FP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

### 3. METODOLOGI PENELITIAN

#### 3.1 Metode Pengumpulan Data

Data yang berasal dari rekam medik Rumah Sakit Siti Khadijah Palembang yang berjumlah 613 *record* dengan atribut sebanyak 9 atribut dalam kurun waktu hampir 3 tahun. Untuk dataset yang di gunakan untuk lebih jelasnya dapat di lihat pada gambar dibawah ini data pasien diabetes berformat pdf.

| No. | Tgl masuk | Tgl keluar | Kadar glukosa puasa(mg/dl) | Kadar glukosa setelah 2 jam(mg/dl) | Tekanan darah (mm/hg) | Berat Badan(Kg) | Umur(Th) | Jenis kelamin(L/P) |
|-----|-----------|------------|----------------------------|------------------------------------|-----------------------|-----------------|----------|--------------------|
| 1   | 03-Jun    | 07-Apr     | 300                        | 262                                | 120/80                | 77              | 52       | P                  |
| 2   | 05-Apr    | 09-Apr     | 271                        | 217                                | 120/80                | 68              | 50       | P                  |
| 3   | 29-Apr    | 02-May     | 251                        | 201                                | 120/70                | 54              | 58       | L                  |
| 4   | 01-Jan    | 04-Jan     | 300                        | 248                                | 120/70                | 60              | 61       | P                  |
| 5   | 21-Mar    | 21-Mar     | 216                        | 178                                | 120/70                | 47              | 45       | P                  |
| 6   | 02-Feb    | 10-Apr     | 218                        | 180                                | 120/80                | 61              | 63       | P                  |
| 7   | 30-Jun    | 30-Jun     | 226                        | 150                                | 120/70                | 55              | 46       | L                  |
| 8   | 23-Mar    | 26-Mar     | 265                        | 215                                | 120/80                | 59              | 56       | L                  |
| 9   | 17-Mar    | 20-Mar     | 290                        | 245                                | 120/70                | 62              | 65       | P                  |
| 10  | 26-Feb    | 26-Feb     | 159                        | 106                                | 120/80                | 55              | 59       | L                  |
| 11  | 03-Jun    | 07-Jun     | 340                        | 284                                | 130/70                | 58              | 49       | L                  |
| 12  | 05-Apr    | 10-Apr     | 290                        | 221                                | 120/80                | 55              | 63       | P                  |
| 13  | 17-Apr    | 21-Apr     | 232                        | 212                                | 120/80                | 64              | 54       | L                  |
| 14  | 01-Jan    | 04-Jan     | 310                        | 254                                | 130/70                | 70              | 62       | L                  |
| 15  | 21-Mar    | 25-Mar     | 255                        | 198                                | 120/70                | 40              | 50       | P                  |

Sumber: rekam medik rs siti khadijah palembang.

**Gambar 2.** Dataset Pasien Diabetes Melitus [2]

#### 3.2 Teknik Analisis Data

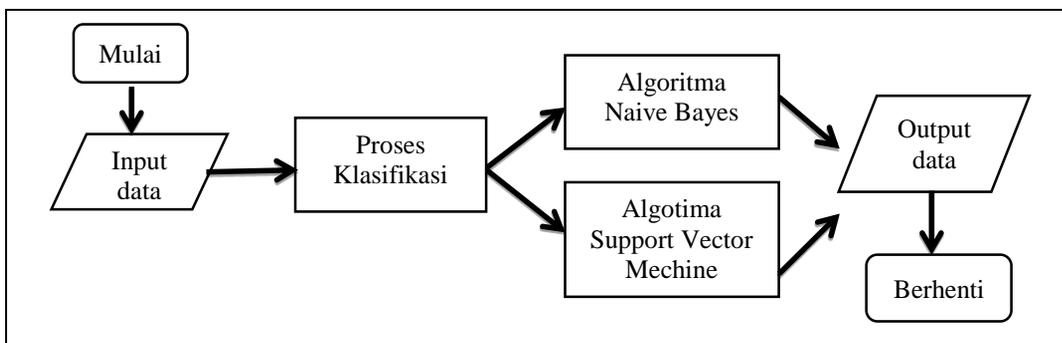
(Dalam et al., n.d.)Dari pengumpulan data yang dijelaskan sebelumnya tidak seluruh data yang diperoleh dapat diproses untuk klasifikasi perlu ada analisis data terlebih dahulu karena adaatribut yang tidak perlu harus melalui berbagai tahapan analisis data agar mendapatkan data yang berkualitas.Ada beberapa tahapan untuk pengolahan data yang bertujuan untuk mendapatkan data yang berkualitas, teknik yang dilakukan untuk proses analisis data sebagai berikut:

- 1) **Selection Data.** Padatahap seleksi data dilakukan penyeleksian data hasil dari tahap *cleaning* data, karena tidak semua dataset yang diperoleh digunakan seluruhnya untuk itu

ada beberapa atribut yang berisi informasi tidak masuk dalam kebutuhan data yang akan diolah seperti atribut tanggal masuk, tanggal keluar akan dihapus dari dataset. Hasil dari *selection* data yang akan menjadi bahan untuk digunakan pada tahap analisis selanjutnya.

- 2) **Cleaning Data.** Untuk tahapan *cleaning* data dilakukan proses mengoreksi data dengan menghilangkan *missing value*. Disebut data *missing value* jika atribut didalam data tidak berisi atau kosong (*null*), serta data yang tidak lengkap. Kemudian data yang akan dibersihkan akan melalui proses *cleaning* dengan melalui cara *Replace Missing Values* untuk mengisi nilai rata-rata atribut tertentu disetiap daerah yang kosong yang mengacu pada atributnya.
- 3) **Transformation Data.** Tahap ini dilakukan proses transformasi atau normalisasi data kedalam format yang dapat dikelola oleh *system*. Dengan cara normalisasi menggunakan tools WEKA dan mengubah *format* data awal sesuai dengan kategori dikarenakan analisis *asosiasi* hanya bisa menerima input data kategorikal, transformasi pada kolom data kontinu dilakukannya proses diskretisasi kemudian dilanjutkan dengan proses binerisasi.

### 3.3 Diagram Alur Penelitian



Gambar 3. Alur Penelitian [3]

## 4. HASIL DAN PEMBAHASAN

Pada tahap pengujian pembagian dataset penyakit diabetes menggunakan *10-fold cross validation*. Yang bertujuan untuk mengetahui hasil akurasi dan menganalisis algoritma *naïve bayes* dan *support vector machine* berdasarkan hasil akurasi dan error yang akan diterapkan pada kernel polynomial dan RBF untuk menghasilkan tingkat akurasi terbaik pada algoritma.

### 4.1 Analisis Klasifikasi Naïve Bayes

Proses tahap analisis klasifikasi untuk algoritma Naïve Bayes dengan hasil *Confussion matrix* sebagai berikut:

**Tabel 2. Hasil Confussion Matrix Naïve Bayes**

| Kelas   | Klasifikasi |         | Total |
|---------|-------------|---------|-------|
|         | Positif     | Negatif |       |
| Positif | 323         | 24      | 347   |
| Negatif | 10          | 72      | 82    |
| Total   | 333         | 96      | 429   |

Berdasarkan pada tabel 2 diatas hasil dari perbandingan pada variabel status positif diperoleh dengan nilai klasifikasi 323 kasus, pada kelas yang masuk pada status negatif tetapi pada klasifikasi masuk kedalam status positif sebanyak 10 kasus. Untuk klasifikasi status negatif yang tepat klasifikasi sebanyak 72 kasus namun, terdapat 24 kasus untuk yang salah klasifikasi oleh

sistem. Total data untuk kelas pada status positif ada 347, sedangkan pada status negatif ada 82. Jumlah untuk data klasifikasi pada status positif ada 333 sedangkan status negatif ada 39. Untuk mengetahui tingkat akurasi dan *error* pada klasifikasi penyakit diabetes melitus pada data *training* dengan algoritma naïve bayes sebagai berikut:

$$Accuracy = \frac{\Sigma \text{klasifikasi tepat}}{\Sigma \text{data training}} = \frac{323+72}{429} = 0.920746 = 92.0746\%$$

$$Error = \frac{\Sigma \text{klasifikasi salah}}{\Sigma \text{data training}} = \frac{10+24}{429} = 0.079254 = 7.9254\%$$

#### 4.2 Analisis Klasifikasi Support Vector Machine Kernel Polynomial

Analisis klasifikasi untuk algoritma *support vector machine* dengan kernel polynomial dengan hasil *Confusion matrix* sebagai berikut:

**Tabel 3. Confusion matrix SVM dengan Kernel Polynomial**

| Kelas   | Klasifikasi |         | Total |
|---------|-------------|---------|-------|
|         | Positif     | Negatif |       |
| Positif | 345         | 2       | 347   |
| Negatif | 14          | 68      | 82    |
| Total   | 359         | 70      | 429   |

Berdasarkan pada tabel 3 merupakan hasil analisis akurasi algoritma *support vector machine* dengan kernel *polynomial* dengan hasil klasifikasi positif sebanyak 345 kasus, untuk klasifikasi positif tetapi masuk kedalam kelas negatif tidak tepat dengan total 14 kasus, selanjutnya klasifikasi negatif tapi masuk kedalam kelas positif ada 2 kasus saja selanjutnya kasus klasifikasi negatif yang tepat sebanyak 68 kasus. Maka didapat nilai akurasi dan *error* sebagai berikut:

$$Accuracy = \frac{\Sigma \text{klasifikasi tepat}}{\Sigma \text{data training}} = \frac{345+68}{429} = 0.962704 = 96.2704\%$$

$$Error = \frac{\Sigma \text{klasifikasi salah}}{\Sigma \text{data training}} = \frac{14+2}{429} = 0.037296 = 3.7296\%$$

#### 4.2 Analisis Klasifikasi Support Vector Machine Kernel RBF

Analisis klasifikasi untuk algoritma *support vector machine* dengan kernel RBF dengan hasil *Confusion matrix* sebagai berikut:

**Tabel 4. Confusion matrix SVM dengan Kernal RBF**

| Kelas   | Klasifikasi |         | Total |
|---------|-------------|---------|-------|
|         | Positif     | Negatif |       |
| Positif | 347         | 0       | 347   |
| Negatif | 82          | 0       | 82    |
| Total   | 429         | 0       | 429   |

Dapat dilihat pada tabel 4 diatas merupakan hasil dari *classification* menggunakan algoritma *support vector machine* dengan kernel RBF. Untuk status positif tepat mendapatkan hasil klasifikasi sebesar 347 kasus, Untuk kelas negatif tetapi masuk kedalam status positif sebanyak 82

kasus sedangkan kasus yang tergolong salah didalam sistem sebanyak 0 kasus serta untuk klasifikasi status negatif yang tepat juga 0 kasus. Berikut diperoleh nilai akurasi dan *error*nya.

$$Accuracy = \frac{\Sigma \text{klasifikasi tepat}}{\Sigma \text{data training}} = \frac{347+0}{429} = 0.808857 = 80.8857\%$$

$$Error = \frac{\Sigma \text{klasifikasi salah}}{\Sigma \text{data training}} = \frac{82+0}{429} = 0.191142 = 19.1142\%$$

### 4.3 Perbandingan Hasil klasifikasi Naïve Bayes dan Support Vector Machine

Perbandingan dari hasil kedua algoritma ini digunakan untuk menentukan metode yang tingkat akurasinya tertinggi. Dari hasil kedua algoritma diperoleh akurasi dapat dilihat pada tabel berikut:

**Tabel 5.** Hasil Perbandingan Algoritma

|         | Naïve Bayes | SVM Kernel Polynomial | SVM Kernel RBF |
|---------|-------------|-----------------------|----------------|
| Akurasi | 92.0746%    | 96.2704%              | 80.8857%       |

Untuk hasil perbandingan antara algoritma naïve bayes dan support vector machine dapat ditarik kesimpulan dengan melihat tabel 5 bahwa nilai yang diperoleh hasil dari pengujian *cross validation 10 folds* dengan hasil akurasi yaitu algoritma SVM dengan Kernel *Polynomial* yang nilai akurasinya tertinggi sebesar 96.2704%.

## 5. KESIMPULAN

Berdasarkan dari pembahasan yang telah diuraikan sebelumnya maka kesimpulan dari penelitian pada perbandingan metode *naïve bayes* dan *support vector machine* untuk klasifikasi penyakit diabetes melitus pada RSI Siti Khadijah Palembang sebagai berikut:

- 1) Dengan menggunakan data *training* dan data *testing* kemudian diukur untuk mengetahui tingkat akurasi dengan evaluasi *confusion matrix* terhadap beberapa percobaan terbaik. Untuk algoritma SVM menggunakan dua kernel yaitu kernel *polynomial* dan RBF menghasilkan nilai akurasi dari algoritma SVM dengan kernel *Polynomial* memiliki nilai akurasi yang tertinggi dengan nilai **96.2704%** bisa dikatakan lebih akurat bila dibandingkan dengan algoritma *naïve bayes* dengan tingkat akurasi **92.0746%**.
- 2) Maka dapat disimpulkan bahwa algoritma yang paling tepat untuk klasifikasi penyakit diabetes melitus pada RSI Siti Khadijah Palembang adalah algoritma *support vector machine*.
- 3) Dari penelitian sebelumnya yang dilakukan oleh Widyawati & Susanto (2019) dapat disimpulkan mengapa ada faktor yang menyebabkan ada salah satu yang hasilnya berbeda yaitu algoritma naïve bayes yang lebih unggul. Hal ini di karenakan peneliti menggunakan tahapan *stopword removal* yang artinya ada kemungkinan merubah isi konteks yang pada dasarnya akan memiliki arti berbeda bahkan berlawanan arah. Berdasarkan analisa juga dapat disimpulkan bahwa dari perhitungan secara manual juga tidak menjamin keakuratan dalam pengklasifikasian data.

Agar penelitian ini lebih baik kedepannya maka penulis memberikan saran agar diharapkan dapat berguna untuk penelitian selanjutnya adapun saran yang diberikan:

- 1) Diharapkan untuk peneliti lainnya dalam membandingkan algoritma yang lain yang lebih baik untuk klasifikasi penyakit diabetes melitus.
- 2) Diharapkan penelitian ini dapat membantu mahasiswa dalam memahami konsep data mining yang berkaitan dengan algoritma yang berkaitan. Membandingkan hasil yang telah dilakukan dengan penambahan kernel dan yang tidak menggunakan kernel.

#### DAFTAR PUSTAKA

- [1] Andriyanto, I., & Santoso, E. (2017). Pemodelan Sistem Pakar Untuk Menentukan Penyakit Diabetes Mellitus Menggunakan Metode Naive Bayes Studi Kasus: Puskesmas Poncokusumo Malang. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(2), 880–887.
- [2] Arifin, O., & Sasongko, T. B. (2018). *ANALISA PERBANDINGAN TINGKAT PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN NAIVE BAYES CLASSIFIER. 2016*, 67–72.
- [3] Dalam, C., Penyakit, P., Di, M., Sakit, R., & Siwi, K. (n.d.). *PERBANDINGAN KINERJA METODE NAIVE BAYES DAN*. 1–8.
- [4] Nugroho, R. A., Tarno, & Prahutama, A. (2017). Klasifikasi Pasien Diabetes Mellitus Menggunakan Metode Smooth Support Vector Machine (Ssvm). *Gaussian*, 6, 439–448.
- [5] Rahman, M. F., Alamsah, D., Darmawidjaja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36. <https://doi.org/10.26555/jifo.v11i1.a5452>
- [6] Riyanto, U. (2018). *ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASIKAN JUMLAH PEMBACA*. 62–72.
- [7] Saleh, A., & Utama, U. P. (2016). *Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. June.
- [8] Sukmawati, Lawi, A., & Thamrin, S. A. (2013). *Klasifikasi Penyakit Diabetes Melitus Tipe II Menggunakan Metode Support Vector Machine*.
- [9] Widyawati, & Sutanto. (2019). *Perbandingan Algoritma Naive Bayes Dan Support Vector Machine ( Svm )*. 3(2), 178–194.