

PENERAPAN ALGORITMA NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION (PSO) DALAM MENANGANI KASUS KANKER PAYUDARA

Wahyu Hadikristanto¹⁾, Vidya Ayu Anisa Fitri²⁾

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa

wahyu.hadikristanto@pelitabangsa.ac.id

Disetujui, 26 September 2018

Abstraksi

Kanker payudara adalah kanker yang menyerang sel-sel pada payudara. Kanker payudara menempati urutan ke-dua penyakit yang mematikan pada wanita. Penyakit ini penderitanya semakin banyak di Negara berkembang karena meningkatnya gaya hidup masyarakat. Gaya hidup yang berkaitan dengan kanker misalnya meroko, aktivitas fisik yang kurang, dan konsumsi obat diet. Berdasarkan riwayat data penderita kanker payudara dapat diprediksi dengan data mining yang dapat membantu para tenaga medis. Data mining adalah suatu proses mengidentifikasi data agar menjadi sebuah informasi maupun keputusan. Penelitian ini menggunakan algoritma naïve bayes berbasis Particle Swarm Optimization (PSO). Hasil akurasi menggunakan algoritma naïve bayes mendapatkan nilai 73,09% dan nilai AUC 0,691. Sedangkan hasil algoritma naïve bayes berbasis PSO mendapatkan nilai 75,50% dan nilai AUC 0,682. Dari penelitian yang dilakukan bahwa naïve bayes berbasis PSO mendapatkan nilai akurasi tertinggi dengan peningkatan sebesar 2,41% dalam memprediksi penyakit kanker payudara.

Kata Kunci : Kanker payudara, Data mining, Algoritma naïve bayes, PSO.

Abstract

Breast cancer is a cancer that attacks cells in the breast. Breast cancer is carried out for two deadly diseases in women. The disease is increasing in developing countries because of people's lifestyle learning. Lifestyle related to cancer, for example, lack of physical activity, and distribute diet drugs. Based on history of cancer patients data can be predicted by data mining that can help medical staff. Data mining is the process of making data so that it becomes information or decision. This study uses naïve bayes algorithm based on Particle Swarm Optimization (PSO). The results of the accuracy using the naïve Bayes algorithm obtained a value of 73.09% and the AUC value of 0.691. While the results of the naïve bayes algorithm based on the PSO get a value of 75.50% and the AUC value of 0.682%. Research conducted with naïve bayes based on PSO, with a contribution rate of 2.41% in predicting breast cancer.

Keywords : Breast cancer, Data mining, Algorithm naïve bayes, PSO

1. Pendahuluan

Tingkat kematian kanker payudara setiap tahunnya mengalami peningkatan karena kurangnya informasi dan kesadaran terhadap penyakit kanker payudara pada penderita. Data Globocan (Global Burden Cancer) menyebutkan di tahun 2018 terdapat 18,1 juta kasus baru dengan angka kematian sebesar 9,6 juta kematian, dimana 1 dari 5 laki-laki dan 1 dari 6 perempuan di dunia mengalami kejadian kanker. Data tersebut juga menyatakan 1 dari 8 laki-laki dan 1 dari 11 perempuan, meninggal karena kanker. angka kejadian untuk perempuan yang tertinggi adalah kanker payudara yaitu sebesar 42,1 per 100.000 penduduk dengan rata-rata kematian 17 per 100.000 penduduk. Berdasarkan data Riskesdas (Riset Kesehatan Dasar), prevalensi tumor/kanker di Indonesia menunjukkan adanya peningkatan dari 1.4 per 1000 penduduk di tahun 2013 menjadi 1,79 per 1000 penduduk pada tahun 2018. Maka dari itu penelitian ini akan membantu untuk memberikan informasi seseorang terjangkit penyakit kanker payudara menggunakan perhitungan data mining.

Data yang diperlukan untuk melakukan penelitian ini sumbernya dari UCI Repository. Dataset yang di gunakan yaitu Breast Cancer dari dataset ini akan dilakukan penelitian menggunakan algoritma naïve bayes optimasi Particle Swarm Optimization (PSO)..

2. Tinjauan Studi

2.1. Data Mining

Data mining merupakan suatu proses untuk mengidentifikasi pola yang memiliki potensi dan berguna untuk mengelola dataset yang besar (Witten & Frank, 2011). Dalam data mining ada 10 algoritma teratas yang paling berpengaruh yang dipilih oleh peneliti dalam komunitas data mining, dimana 6 (enam)

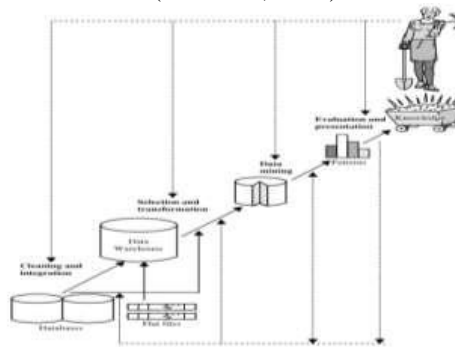
diantaranya adalah algoritma klasifikasi yaitu C4.5, Support Vector Machines (SVM), AdaBoost, k Nearest Neighbor (KNN), Naïve Bayes dan CART (Fayyad, Piatetsky-Shapiro, & Smyth, 1996)

Ilmu data mining seringkali disebut sebagai suatu kegiatan dalam mencari informasi yang terpendam dari suatu data sehingga dapat menjadi sebuah pengetahuan baru untuk digunakan sebagai informasi maupun keputusan (Enri, 2018) Data mining juga dapat diartikan sesuatu hal yang layaknya berharga seperti kata awalnya yakni menambang biasanya menambang identik dengan mencari sesuatu benda ataupun hal yang berharga didalam tempat yang tersembunyi untuk menemukan sesuatu yang berharga seperti emas, berlian dan batu bara maka data mining dapat diartikan sebagai sesuatu kegiatan menambang data untuk mencari pengetahuan yang terpendam dari data yang ada. (Han, Kamber, & Pei, 2012).

2.2. Sistem Penunjang Keputusan

Sistem pendukung keputusan dapat didefinisikan sebagai sebuah sistem yang dimaksudkan untuk mendukung para pengambil keputusan manajerial dalam situasi keputusan tidak terstruktur. SPK dimaksudkan untuk menjadi alat bantu bagi para pengambil keputusan untuk memperluas kapabilitas mereka, namun tidak untuk menggantikan penilaian mereka. Selain itu juga sistem pendukung keputusan ditujukan untuk keputusan – keputusan yang memerlukan penilaian atau pada keputusan keputusan yang sama sekali tidak dapat didukung oleh algoritma(Surya, 2018).

Sistem Pendukung Keputusan adalah sistem penghasil informasi spesifik yang ditujukan untuk memecahkan suatu masalah tertentu yang harus dipecahkan oleh manager pada berbagai tingkatan. Dengan kata lain Sistem Pendukung Keputusan adalah suatu sistem informasi berbasis komputer yang menghasilkan berbagai alternatif keputusan untuk membantu manajemen dalam menangani berbagai permasalahan yang terstruktur dengan menggunakan data dan model(Windarto, 2017).



Gambar 1. Proses Tahapan *Data Mining*

2.3. Naïve Bayes

Algoritma naïve bayes adalah salah satu algoritma yang tergolong kedalam statistical classifier algoritma tersebut pertama kali diperkenalkan oleh Thomas Bayes dimana algoritma ini merupakan adaptasi dari theorem bayes. Algoritma ini mengandalkan sebuah peluang kemungkinan suatu objek.(Han et al., 2012)

Rumus dari theorem bayes sendiri ialah :

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

Keterangan :

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data x merupakan suatu kelas spesifik
- P(H) : Probabilitas hipotesis H (prior probability)
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)
- P(X|H) : Probabilitas X berdasar kondisi pada hipotesis H
- P (X): Probabilitas X

Rumus diatas disesuaikan sebagai berikut :

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)}$$

Rumus Diatas disederhanakan lagi ke dalam bentuk seperti dibawah ini

$$Posterior = \frac{Prior \times Likelihood}{Evidance}$$

Evidance selalu bernilai sama atau tetap pada kelas dalam sebuah sample. Kemudian nilai posterior tersebut dibandingkan dengan nilai posterior kelas lain untuk menentukan kelas klasifikasi dari sebuah

sample. (c.agustina) Berdasarkan persamaan theorem bayes maka inilah rumus naïve bayes classifier.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan :

- X : Data dengan kelas yang belum diketahui
- C_i : Hipotesis data x merupakan suatu kelas spesifik
- P(C_i|X) : Probabilitas hipotesis C_i berdasarkan kondisi X (posteriori probability)
- P(C_i) : Probabilitas hipotesis C_i (prior probabily)
- P(X|C_i) : Probabilitas X berdasar kondisi pada hipotesis C_i
- P(X) : Probabilitas X

Naïve Bayes adalah algoritma klasifikasi probabilitas sederhana yang berdasarkan pada teorema Bayes, asumsi bebas yang kuat (naive), dan model fitur independen (Khadafy & Wahono, 2015).

2.4. Algoritma Metaheuristic

Algoritma Metaheuristic adalah metode pengoptimalan yang dapat memecahkan masalah dengan menjelajahi ruang pencarian solusi. Namun, kinerja algoritme ini sangat bergantung pada pengaturan parameter mereka, dan tidak mudah untuk menetakannya secara akurat serta sepenuhnya bergantung pada karakteristik masalah. Untuk menyempurnakan parameter secara otomatis, banyak metode telah diajukan untuk mengatasi tantangan ini contohnya particle swam optimization (PSO), algoritma Gentik, ant colony optimization (ACO) dll (Ayumi, Rere, Fanany, & Arymurthy, 2017). Meteheuristic memiliki beberapa karakteristik dasar yaitu :

1. Metaheuristik adalah strategi yang memandu proses pencarian
2. Tujuan dari metaheuristik adalah untuk menjelajahi ruang pencarian secara efficient untuk menemukan solusi optimal.
3. Teknik metaheuristik berkisar dari prosedur pencarian local yang sederhana sampai proses pembelajaran yang kompleks
4. Meteheuristik adalah metode pendekatan dan biasanya non-deterministik
5. Metaheuristik dapat terdiri dari penggabungan beberapa mekanisme supaya proses pencarian tidak terjebak dalam daerah terbatas di ruang pencarian.
6. Konsep dasar dari metaheuristik memungkinkan pendeskripsian secara abstrak
7. Metaheuristik bersifat general/umum sehingga dapat diterapkan dalam berbagai macam persoalan
8. Metaheuristik dapat menggunakan domain pengetahuan khusus dalam bentuk heuristik yang dikendalikan dengan strategi tingkat lanjut.

Metaheuristik dapat menggunakan pengalaman yang didapat selama proses pencarian untuk menuntun proses pencarian.

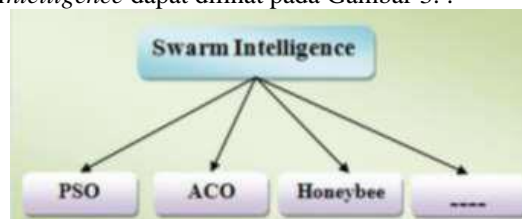
2.5. Swarma Intelligence

Swarm Intelligence adalah disiplin keilmuan dari sistem cerdas yang berhubungan dengan sistem alami dan buatan, yang terdiri dari banyak individu (populasi) yang berkoordinasi menggunakan konsep kontrol desentralisasi (kecerdasan sosial dalam berkelompok) dan self organized (kecerdasan personal/ terorganisir secara mandiri). Swarm Intelligence secara singkat bisa juga disebut sebagai kecerdasan berkelompok (Cholissodin & Riyandani, 2016). Secara umum konsep swarm intelligence dapat di- petakan seperti Gambar 2.



Gambar 2. Konsep Swarm Intelligence

Macam – macam Swarm Intelligence dapat dilihat pada Gambar 3. :



Gambar 3. Jenis Swarm Intelligence

Karakteristik dari Swarm Intelligence:

1. Mengharuskan untuk membuat desain ulang representasi solusi yang sesuai untuk setiap goal atau constraint maupun kasus berbeda.
2. Terkadang tidak mencapai solusi global optimal atau mengalami kegagalan (terjebak pada local optimal, konvergensi dini).
3. Beberapa mekanisme alami belum dapat dipahami dengan baik.
4. Mengharuskan untuk setiap problem optimasi yang akan diselesaikan telah terdefinisi dengan baik, mulai dari input datanya apa saja, bagaimana representasi solusinya, constraints-nya apa saja, bagaimana rancangan rumus fitnessnya atau costnya.

Representasi solusi merupakan vektor atau matrik sederhana maupun kompleks dalam bentuk tertentu dan spesifik, yang dapat berbeda bentuknya atau diubah secara dinamis sesuai dengan cara pandang seseorang terkait pemahamannya terhadap problem yang akan diselesaikan. (Cholissodin & Riyandani, 2016)

2.6. Particle Swarm Optimization

Particle swarm optimization, disingkat sebagai PSO, didasarkan pada perilaku sebuah kawanan serangga, seperti semut, rayap, lebah atau burung. Algoritma PSO meniru perilaku sosial organisme ini (Cholissodin & Riyandani, 2016). Perilaku sosial terdiri dari tindakan individu dan pengaruh dari individu-individu lain dalam suatu kelompok. (Cholissodin & Riyandani, 2016)

Kata "partikel" menunjukkan individu, misalnya seekor burung dalam kawanan burung. Setiap individu atau partikel berperilaku saling terhubung dengan 21 cara menggunakan kecerdasannya (intelligence) sendiri dan juga dipengaruhi perilaku kelompok kolektifnya. (Cholissodin & Riyandani, 2016) Dengan demikian, jika satu partikel atau seekor burung menemukan jalan yang tepat atau pendek menuju ke sumber makanan, sisa kelompok yang lain juga akan dapat segera mengikuti jalan tersebut meskipun lokasi mereka jauh dikelompok tersebut. (Cholissodin & Riyandani, 2016)

Metode optimasi yang didasarkan pada swarm intelligence ini disebut algoritma behaviorally inspired sebagai alternatif dari algoritma genetika, yang sering disebut evolution based procedures. Algoritma PSO ini awalnya diusulkan oleh J. Kennedy dan R. C. Eberhart. (Cholissodin & Riyandani, 2016) Dalam konteks optimasi multi variabel, kawanan diasumsikan mempunyai ukuran tertentu atau tetap dengan setiap partikel posisi awalnya terletak di suatu lokasi yang acak dalam ruang multidimensi. Setiap partikel diasumsikan memiliki dua karakteristik posisi dan kecepatan. (Cholissodin & Riyandani, 2016) PSO memiliki tiga komponen utama diantaranya: partikel, komponen kognitif dan komponen sosial, serta kecepatan partikel. Dan setiap partikel merepresentasikan solusi penyelesaian. Pembelajaran partikel terdiri dari dua faktor yaitu pengalaman partikel (disebut cognitive learning) dan kombinasi pembelajaran dari keseluruhan swarm (disebut social learning). (Cholissodin & Riyandani, 2016)

Terdapat beberapa komponen dalam Algoritma PSO diantaranya adalah sebagai berikut:

1. Swarm, merupakan jumlah partikel dalam populasi pada suatu algoritma. Ukuran swarm bergantung pada seberapa kompleks masalah yang dihadapi. Secara umum, ukuran swarm pada algoritma PSO cenderung lebih kecil jika dibandingkan dengan algoritma evolusioner yang lain dalam mencari solusi terbaik.
2. Partikel, merupakan individu dalam suatu swarm yang merepresentasikan solusi penyelesaian masalah. Setiap partikel memiliki posisi dan kecepatan yang ditentukan oleh representasi solusi pada saat itu.
3. Personal best (pBest), merupakan posisi terbaik yang pernah dicapai partikel dengan membandingkan fitness pada posisi partikel sekarang dengan sebelumnya. Personal best dipersiapkan untuk mendapatkan solusi terbaik.
4. Global best (gBest), merupakan posisi terbaik partikel yang diperoleh dengan membandingkan nilai fitness terbaik dari keseluruhan partikel dalam swarm.
5. Kecepatan (velocity), v merupakan vektor yang menentukan arah perpindahan posisi partikel. Perubahan velocity dilakukan setiap iterasi dengan tujuan memperbaiki posisi partikel semula.
6. Bobot inersia (inertia weight), w digunakan untuk mengontrol dampak dari perubahan velocity yang diberikan oleh partikel.
7. Koefisien akselerasi, merupakan faktor pengontrol sejauh mana partikel berpindah dalam satu iterasi. Secara umum nilai koefisien akselerasi $C1$ dan $C2$ adalah sama yaitu dalam rentang 0 sampai 4. Namun demikian, nilai tersebut dapat ditentukan sendiri untuk setiap penelitian berbeda. Particle Swarm Optimization (PSO) sering digunakan dalam penelitian, karena PSO memiliki kesamaan sifat dengan Genetic Algorithm (GA) (Widiastuti et al., 2014).

Keuntungan dari PSO adalah mudah diterapkan dan ada beberapa parameter untuk menyesuaikan. Sistem PSO diinisiasi oleh sebuah populasi solusi acak dan selanjutnya mencari titik optimum dengan cara memperbarui tiap hasil pembangkitan. Pendekatan yang digunakan lebih sistematis matematika untuk menemukan solusi.

2.7. K-Fold Cross Validation

Validasi dilakukan untuk menguji model algoritma yang digunakan. K-folds cross validation merupakan metode untuk mengetahui tingkat keberhasilan pada model algoritma dengan cara melakukan pengujian ulang atribut input yang acak, dalam metode ini data dibagi menjadi k subset secara acak, satu subset digunakan untuk data testing dan sisanya untuk data training, (Banjarsari, Budiman, & Farmadi, 2016).

10 Fold Cross Validation merupakan salah satu metode pengujian yang digunakan untuk pembelajaran terlatih (supervised learning). Setiap fold dibagi menjadi beberapa subset dengan ukuran sama pada tiap subsetnya. Kemudian akan dilakukan training sebanyak 10 kali dengan menggunakan 9 fold untuk training set dan 1 fold digunakan sebagai test set (Defiyanti & Kom, 2013).

Cara kerja k-folds cross validation, yaitu total data dibagi menjadi n bagian, iterasi atau fold ke 1, yaitu bagian ke 1 menjadi testing, bagian sisanya menjadi data training, kemudian hitung akurasi menggunakan persamaan berikut :

$$Accuracy = \frac{\text{jumah klasifikasi benar}}{\text{jumlah data uji}} \times 100\%$$

Keterangan : Jumlah klasifikasi benar : jumlah prediksi klasifikasi yang tepat Jumlah data uji : jumlah dataset yang digunakan untuk testing

2.8. Kanker Payudara

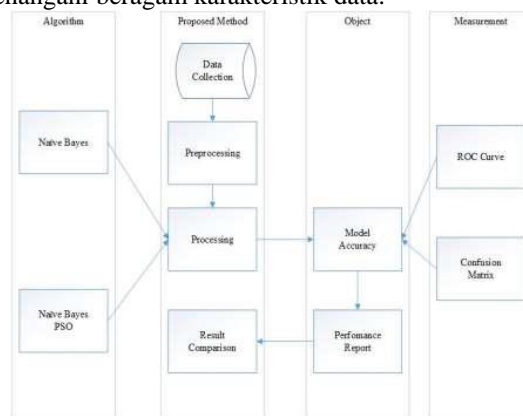
Kanker merupakan salah satu penyebab kematian baik di Negara maju maupun Negara berkembang. Penyakit kanker penderitanya semakin banyak di Negara berkembang karena meningkatnya gaya hidup masyarakatnya. Gaya hidup yang berkaitan dengan kanker misalnya meroko, aktivitas fisik yang kurang, dan konsumsi obat diet. Kanker yang sering dijumpai pada wanita di Indonesia adalah kanker serviks, kanker payudara, kanker paru, kanker kolorektal, dan kanker hati. (Yuliana, 2018)

Penyebab timbulnya kanker payudara belum diketahui secara pasti. Faktor utama yang diduga berperan dalam proses kejadian kanker payudara adalah hormon estrogen, tetapi bagaimana mekanismenya belum jelas. Hormon estrogen dapat meningkatkan proses proliferasi dan pertumbuhan sel-sel spesifik pada tubuh serta bertanggung jawab terhadap sebagian besar sifat seksual sekunder wanita. Pada payudara estrogen dapat menyebabkan pengendapan lemak dalam kelenjar payudara. (Nasution, Siregar, & Asfriyanti, 2018)

Gejala kanker payudara pada umumnya tidak menimbulkan gejala jika ukurannya masih kecil. Jika kanker payudara menimbulkan benjolan yang dapat dirasakan dengan cara diraba, pasien hanya sering mengeluhkan benjolan namun tidak merasakan nyeri pada benjolan. (Yuliana, 2018)

3. Kerangka Konsep

Objek yang akan diteliti ialah model akurasi dari hasil penerapan algoritma yang berbeda yakni dengan menggunakan algoritma NB (naïve bayes) atau NB PSO (naïve bayes berbasis particle swarm intelligence) serta melihat sejauh mana keefektifan algoritma NB PSO dalam menangani beragam data yang berbeda serta merubah dua feature selection yakni population size dan inertia weight dalam menentukan keakuratan yang tertinggi. Untuk dapat melihat hal tersebut digunakan dua metode evaluasi yakni menggunakan confusion matrix dan ROC Curve. Setelah hal tersebut akan terlihat masing-masing laporan dari tiap hasil pengujian, hasil akhir dari penelitian ini ialah dengan menggabungkan semua laporan dari tiap pengujian kedalam satu tabel untuk dapat diamati secara detail seberapa besar keefektifan metode optimasi menggunakan PSO dalam menangani beragam karakteristik data.



Gambar 4. Kerangka Berpikir

4. Design Penelitian/Metodologi

4.1. Tahapan Analisa Data

Datase kanker payudara yang berasal dari UCI Respository mempunyai jumlah dataset berjumlah 286

dan data terdiri dari 9 atribut dan 1 atribut sebagai class. Atribut class mempunyai dua nilai yaitu no-recurrence-events dan recurrence-events. Jumlah nilai no-recurrence-events 201 sedangkan nilai recurrence-events 85 data. Dalam membuat sebuah keputusan atau menentukan akurasi perlu adanya data yang berkualitas untuk itu dilakukan tahap data preprocessing. Data preprocessing digunakan untuk membersihkan data dari missing value, ketidak konsistenan, data tidak lengkap, dan noise data. Berikut tahap-tahap data preprocessing antara lain, yaitu :

1. *Data Cleaning*

Data cleaning adalah pembersihan data yang tidak dibutuhkan karena adanya kesalahan data. Pembersihan data dilakukan jika ada missing data, missing data adalah data yang hilang. Pada dataset kanker payudara terdapat missing value yaitu:

Tabel 1. Missing Value

No	Atribut	Jumlah <i>Missing Value</i>
1	<i>Node Caps</i>	8
2	<i>Breast Quad</i>	1

Berdasarkan table di atas atribut yang terdapat missing value ada 2 yaitu atribut *node_caps* terdapat 8 data yang kosong dan atribut *Breast_Quad* terdapat 1 data yang kosong. Data yang kosong tersebut kemudian dilakukan menggunakan teknik tuple atau record data yang bernilai kosong menggunakan filter example pada aplikasi RapidMiner.

2. *Data Reduction*

Data reduction merupakan tahap data preprocessing dengan mengurangi jumlah data yang dihasilkan dari data asli, pengurangan tersebut dapat dilakukan dengan cara pengurangan jumlah atribut, namun tetap memilih atribut yang penting yang akan digunakan dalam teknik data mining.

Dataset yang terdiri dari 9 atribut dilakukan proses data preprocessing dengan tahap reduction untuk mengurangi atribut. Factor penyebab terjadinya penyakit payudara adalah umur, menopause, kelenjar getah bening, dan ukuran tumor, maka dari 10 atribut yang ada dalam dataset akan dikurangi 1 menjadi 9 atribut yang akan digunakan. Berikut table yang akan digunakan :

Tabel 2. Atribut yang Digunakan

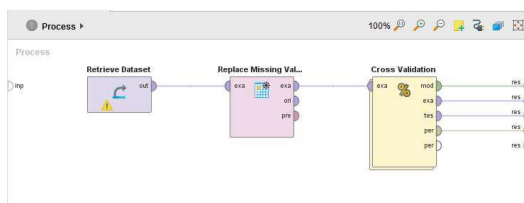
No	Atribut	Keterangan
1	<i>Class</i>	Label <i>no recurrence event</i> atau <i>recurrence event</i>
2	<i>Age</i>	usia pasien pada saat diagnosis
3	<i>Menopause</i>	apakah pasien pra atau pasca menopause pada saat diagnosis
4	<i>Inv_Nodes</i>	jumlah kelenjar getah bening yang mengandung kanker payudara metastatik terlihat pada pemeriksaan histologis
5	<i>Node_Caps</i>	kanker yang tidak bermetastasis ke kelenjar meskipun di luar penyakit tumor tetap terkanung oleh kapsul kelenjar getah bening
6	<i>Deg_Malig</i>	jika kanker tidak bermetastasis ke kelenjar getah bening, meskipun di luar lokasi asli tumor itu mungkin tetap "terkandung" oleh kapsul kelenjar getah bening. Namun, seiring waktu, dan dengan penyakit yang lebih agresif, tumor dapat menggantikan kelenjar getah bening dan kemudian menembus kapsul, memungkinkannya untuk menyerang jaringan sekitarnya.
7	<i>Breast</i>	Kanker payudara jelas dapat terjadi pada payudara yang baik
8	<i>Breast_Quad</i>	payudara dapat dibagi menjadi empat kuadran, menggunakan puting sebagai titik sentral
9	<i>Irradiat</i>	terapi radiasi adalah pengobatan yang menggunakan energi x tinggi

5. **Hasil Penelitian Dan Pembahasan**

Berikut adalah hasil pengujian yang dilakukan:

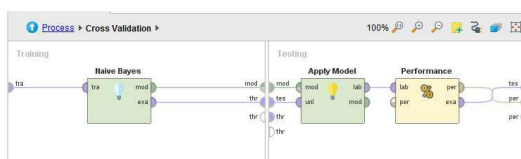
5.1. **Pengujian 1 (Naive Bayes)**

Pengujian pertama dilakukan dengan menggunakan Rapidminer 9.0.2 untuk melakukan preprocessing mode digunakan menggunakan algoritma klasifikasi yaitu algoritma naive bayes. Berikut adalah gambar model dari pengujian pertama pada Gambar 1 :



Gambar 5. Pengujian 1

Gambar 1 menggambarkan rangkaian model pengujian 1 secara global dimana pada retrieve dataset nantinya akan dimasukan dataset kanker payudara. Kemudian untuk mencari data yang kosong menggunakan missing value kemudian untuk memvalidasi model dari algoritma naïve bayes digunakan metode cross validation. Dimana didalamnya terdapat performance dengan menggunakan confusion matrix dan roc curve sebagai model evaluasi dari kinerja algoritma naïve bayes. Untuk lebih jelasnya dapat dilihat pada gambar 2 berikut :



Gambar 6. Sub Proses Cross Validation

Terakhir dibagian testing terdapat fitur performance dimana fitur tersebut digunakan untuk mengevaluasi hasil kinerja algoritma naïve bayes dengan parameter pengukuran confusion matrix (accuracy, recall, precision) dan ROC Curve.

Pada pengujian pertama ini data uji untuk diterapkan pada algoritma naïve bayes tanpa menggunakan metode optimasi. Sebelum diproses data sudah lebih dahulu memasuki tahap preprocessing sehingga data yang akan diproses sudah bersih dan dapat digunakan untuk proses klarifikasi. Hasil yang didapatkan dari pengujian pertama ini menggunakan algoritma naïve bayes adalah sebagai berikut :

1. Accuracy

Hasil accuracy menggunakan algoritma naïve bayes yaitu 73.09%, untuk lebih jelasnya dapat dilihat pada gambar 3 sebagai berikut :

accuracy: 75.50% +/- 6.10% (micro average: 75.52%)

	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	180	49	78.60%
pred. recurrence-events	21	36	63.16%
class recall	89.55%	42.35%	

Gambar 7. Hasil Accuracy Algoritma Naïve Bayes

2. Precision

Hasil Precision menggunakan algoritma naïve bayes yaitu 59.64%., untuk lebih jelasnya dapat dilihat pada gambar 4 sebagai berikut :

precision: 59.64% +/- 21.06% (micro average: 55.56%) (positive class: recurrence-events)

	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	169	45	78.97%
pred. recurrence-events	32	40	55.56%
class recall	84.08%	47.06%	

Gambar 8. Hasil Precision Algoritma Naïve Bayes

3. Recall

Hasil recall menggunakan algoritma naïve bayes yaitu 47.08%, untuk lebih jelasnya dapat dilihat pada gambar 5 sebagai berikut :

recall: 47.08% +/- 19.44% (micro average: 47.06%) (positive class: recurrence-events)

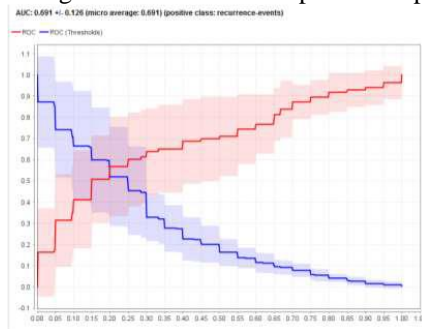
	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	169	45	78.97%
pred. recurrence-events	32	40	55.56%
class recall	84.08%	47.06%	

Gambar 9. Hasil Recall Algoritma Naïve Bayes

4. AUC

Kemudian untuk nilai AUCnya didapatkan nilai 0,691 dengan kategori sebagai excellent

classification. Berikut ini gambar kurva ROC dapat dilihat pada gambar 6 sebagai berikut :



Gambar 10. Hasil AUCI Algoritma Naïve Bayes

5.2. Pengujian 2 Menggunakan PSO

1. Pembobotan PSO

Hasil pembobotan yang didapatkan dari algoritma naïve bayes metode PSO dapat dilihat pada gambar 7 sebagai berikut :

attribute	weight
Age	1
menopa...	0
Ts	0
Ins	0.224
Ncs	0
Dma	0.230
breast	0
Brq	0.119
irradiat	0

Gambar 11. Pembobotan PSO

2. Accuracy

accuracy: 75.50% +/- 6.10% (micro average: 75.52%)

	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	180	49	78.60%
pred. recurrence-events	21	36	63.16%
class recall	89.55%	42.35%	

Gambar 12. Hasil Accuracy Algoritma Naïve Bayes Metode PSO

3. Precision

precision: 62.15% +/- 15.31% (micro average: 63.16%) (positive class: recurrence-events)

	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	180	49	78.60%
pred. recurrence-events	21	36	63.16%
class recall	89.55%	42.35%	

Gambar 13. Hasil Precision Algoritma Naïve Bayes Metode PSO

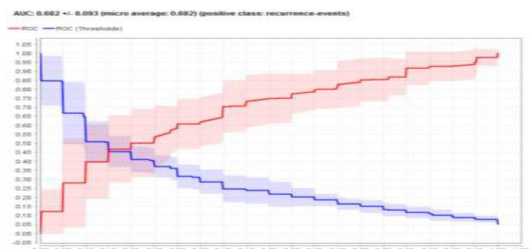
4. Recall

recall: 42.08% +/- 16.91% (micro average: 42.35%) (positive class: recurrence-events)

	true no-recurrence-events	true recurrence-events	class precision
pred. no-recurrence-events	180	49	78.60%
pred. recurrence-events	21	36	63.16%
class recall	89.55%	42.35%	

Gambar 14. Hasil Recall Algoritma Naïve Bayes Metode PSO

5. AUC



Gambar 15. Hasil AUCI Algoritma Naïve Bayes Metode PSO

5.3. Pembahasan

Penelitian yang dilakukan terbagi menjadi dua pengujian, yaitu pengujian menggunakan algoritma naïve bayes dan pengujian kedua menggunakan naïve bayes metode PSO. Hasil dari dua penelitian bisa dilihat di table 1 sebagai berikut :

Tabel 3. Tabel Hasil Penelitian

Pengukuran	Perhitungan	
	<i>Naïve Bayes</i>	<i>Naïve Bayes Metode PSO</i>
<i>Accuracy</i>	73.09%	75.50%
<i>Recall</i>	59.64%	62.15%
<i>Precision</i>	47.08%	42.08%
<i>AUC</i>	0.691	0.682

Berdasarkan hasil yang diperoleh bahwa adanya peningkatan nilai accuracy, recall, precision, dan nilai AUC. Pada algoritma naïve bayes menggunakan metode PSO mendapatkan hasil yang meningkat dibandingkan jika hanya menggunakan satu teknik pembelajaran saja.

Penelitian mengenai penyakit kanker payudara menggunakan dataset UCI Repository dengan menggunakan algoritma naïve bayes metode PSO mengalami peningkatan akurasi yaitu sebesar 2.41%. hal ini menunjukkan bahwa metode Particle Swarm Optimization (PSO) dapat menghasilkan peningkatan akurasi lebih tinggi dibandingkan dengan perhitungan algoritma naïve bayes.

6. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut:

1. Berdasarkan penelitian yang sudah dilakukan algoritma naïve bayes metode PSO mengalami peningkatan nilai sebesar 2.41%. Maka algoritma naïve bayes metode PSO efektif dalam meningkatkan performa algoritma naïve bayes.
2. Berdasarkan hasil pengujian untuk memprediksi kanker payudara dengan menggunakan metode naïve bayes berbasis PSO, data yang digunakan yaitu ada 10 atribut namun yang dipakai dalam penelitian hanya menggunakan 9 atribut.
3. Hasil penelitian yang dilakukan Algoritma naïve bayes mendapatkan hasil akurasi 73,09% sedangkan algoritma naïve bayes metode PSO mendapatkan hasil akurasi 75,50%. Maka nilai akurasi tertinggi yaitu Algoritma Naive bayes metode PSO.

Daftar Pustaka

- Agustina, C. (2018). Optimasi Naive Bayes Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Deteksi Autisme Spectrum Disorder. *Journal Speed Sentra Penelitian Engineering Dan Edukasi*, 10(2),10–14.
- Aliyah Hidayatul Muslimin 2 Kecamatan Sungai Raya Kabupaten Kubu Raya Tahun 2017. *Jurnal Kebidanan*, 7(1), 20–25.
- Ayumi, V., Rere, M. R., Fanany, M. I., & Arymurthy, A. M. (2017). Random Adjustment - Based Chaotic Metaheuristic Algorithms For Image Contrast Enhancement. *Jurnal Ilmu Komputer Dan Informasi*, 13(3), 1576–1580.
- Cholissodin, I., & Riyandani, E. (2016). *Swarm Intelligence(Teori & Case Study)*.
- Enri, U. (2018). Optimasi Parameter Support Vector Machines Untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika Serikat. *Jurnal Gerbang*, 8(1), 65–72.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining To Knowledge Discovery In Databases. *AI Magazine*, 37–54. <https://doi.org/10.1145/240455.240463>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining*.
- Hermawanti, L. (2012). Penerapan Algoritma Klasifikasi C4.5 Untuk Diagnosis Penyakit Kanker Payudara. *Jurnal Teknik Unisfat*, 7(1), 57–64.
- Nasution, W. M., Siregar, F. A., & Asfriyanti. (2018). Pengaruh Pemakaian Kontrasepsi Hormonal Dan Riwayat Keluarga Terhadap Kejadian Kanker Payudara Di Rsud Dr . Pirngadi Medan Tahun 2017. *Jurnal Medika Respati*, 13(2), 39–47. 55
- Pebrianti, D., & Alexander. (2017). Evaluasi Pengetahuan Sebelum Dan Sesudah Penyuluhan Tentang Kanker Payudara Dan Praktek Sadari Di Madrasah
- Prakash, J., Gupta, D. K., & Kumar, R. (2017). Research Article Soft Computing Based Cluster-Head Selection In Mobile Ad-Hoc Network Jay Prakash. *Journal of Artificial Intelligence*, 10, 98–111. <https://doi.org/10.3923/jai.2017.98.111>
- Rinawati. (2017). Penentuan Penilaian Kredit Menggunakan Metode Naive Bayes Berbasis Particle Swarm. *Jurnal Sains Komputer Dan Informatika*, 1(1), 48–58.
- Sari, R. M. (2015). Hubungan Pengetahuan Ibu Dengan Upaya Deteksi Dini Kanker Payudara Melalui Sadari

- Di Kelurahan Nglames Kabupaten Madiun. *Jurnal Ners Dan Kebidanan*, 2(3), 276–281.
<https://doi.org/10.26699/jnk.v2i3.ART.p276-281>
- Septiani, W. D. (2017). Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76–84.
- Taufik, A. (2017). Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Komputer*, III(2), 40–47.
- Via, Y. V., Nugroho, B., & Syafrizal, A. (2015). Sistem Pendukung Keputusan Klasifikasi Tingkat Keganasan Kanker Payudara Dengan Metode Naive.SCAN, X, 63–68.56
- Wibowo, K., Sfenrianto, & Naingolan, K. (2015). Klasifikasi Gangguan Motorik Kasar Anak Menggunakan Naive Bayes Serta Optimasi Dengan Pso Dan Adaboost. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 1(1), 1–10.
- Widiastuti, N. A., Santosa, S., & Supriyanto, C. (2014). Algoritma Klasifikasi Data Mining Naïve Bayes Berbasis Particle Swarm Optimization Untuk Diteksi Penyakit Jantung. *Jurnal Pseudocode*, 1(1), 11–14.
- Witten, I. H., & Frank, E. (2011). *Data Mining: Practical Machine Learning Tools And Techniques*. Elsevier, San Francisco, 3 edition, 629.<https://doi.org/0120884070>, 9780120884070
- Yani, A. (2018). Pemanfaatan Teknologi Dalam Bidang Kesehatan Masyarakat. *Jurnal Kesehatan Masyarakat*, 8(1), 97–102.
- Yuliana. (2018). Risiko dan Deteksi Dini Kanker Payudara. *CDK-261*, 45(2), 144–149.