

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Penelitian ini dikembangkan dari beberapa referensi yang mempunyai keterkaitan dengan metode dan objek penelitian. Penggunaan referensi ini ditujukan untuk memberikan batasan-batasan terhadap metode dan sistem yang nantinya akan dikembangkan lebih lanjut. Berikut adalah hasil dari penelitian sebelumnya.

Penelitian pertama yaitu optimasi metode Naive Bayes menggunakan Particle Swarm Optimization dalam mendeteksi *Autisme Spectrum Disorder* oleh (Agustina, 2017) membahas tentang bagaimana merancang tools yang membantu pakar untuk mengdiagnosa seseorang menderita *Autisme Spectrum Disorder* atau tidak, berdasarkan kriteria pada *Diagnostic and Statistical Manual of Mental Disorder* (DSM-5) yang dikeluarkan oleh APA (*American Psychiatric Assosiasi*). Pada penelitian ini data set diperoleh dari UCI Repository berjumlah 704 record dengan memiliki 20 atribut (10 atribut data perilaku dan 10 karakteristik seseorang). Hasil akhir dari penelitian tersebut yaitu penerapan Particle Swarm Optimization terbukti meningkatkan akurasi pada Algoritma Naive Bayes dengan selisih persentase sebesar 5,68% .

Penelitian kedua yaitu tentang peningkatan akurasi pada produk *furniture* dengan menggunakan algoritma Naive Bayes yang dioptimasi Particle Swarm Optimization oleh (Aulianita & Rifai, 2018). Dalam penelitian ini penerapan metode algoritma tersebut dalam mengklasifikasikan *review* positif dan negatif dimana akan dimana memudahkan pembeli mendapatkan informasi secara efektif sebelum memilih produk yang tepat untuk dibeli. Pengolahan data uji training

menggunakan 100 data review *furniture* positif dan 100 data review *furniture* negatif. Hasil akhir penelitian yaitu akurasi yang telah dilakukan pada metode Naive Bayes *review furniture* yaitu 51,50% dan AUC sebesar 0,5 sedangkan pada saat penambahan Particle Swarm Optimization pada dataset training Naive Bayes dihasilkan akurasi sebesar 93,50% dan AUC sebesar 0,9. Hal ini membuktikan bahwa PSO merupakan salah satu teknik optimasi untuk seleksi fitur dan penambahan bobot atau meningkatkan akurasi.

Penelitian yang ketiga yaitu pemodelan sistem pakar diagnosa penyakit ginjal kronis dengan menggunakan dataset yang diambil dari UCI Repository oleh (Arifin & Ariesta, 2019). Metode yang diterapkan yaitu naive bayes, algoritma dipilih karena cocok diterapkan pada data dengan jumlah yang banyak. Tetapi metode klasifikasi naive bayes memiliki kekurangan dimana probabilitasnya tidak bisa menentukan seberapa akurat nilai akurasi sebuah klasifikasi sehingga dilakukan pengoptimalan dengan cara memberikan nilai bobot pada setiap atribut. Hasil perhitungan dengan Naive Bayes Classification yang diterapkan pada dataset penyakit ginjal kronis menghasilkan nilai akurasi *confusion matrix* sebesar 97.00% dan AUC sebesar 99.8% dan setelah ditambahkan Particle Swarm Optimization untuk pembobotan atribut pada dataset penyakit ginjal kronis hasil akurasi dari Naive Bayes Classification meningkat menjadi 98,75% dan AUC sebesar 99%, sehingga pemodelan sistem pada penelitian ini dapat digunakan oleh ahli Patologi dalam membuat program untuk memprediksi penyakit ginjal kronis.

Penelitian keempat yaitu sistem diagnosis penyakit hepatitis menggunakan metode naive bayes berdasarkan masukan dari gejala-gejala penyakit hepatitis oleh (Koeswara et al., 2020). Dalam penelitian ini mampu melakukan diagnosis penyakit

hepatitis dan didapatkan hasil pengujian dengan akurasi Naive Bayes senilai 84,85%, sedangkan untuk nilai akurasi Optimasi Naive Bayes menggunakan PSO sebesar 92,50% sehingga tampak selisih nilai akurasi yaitu sebesar 7,65%. Sedangkan evaluasi menggunakan ROC curve diperoleh hasil untuk algoritma Naive Bayes bernilai 0,894 dengan tingkat diagnosa *Good Classification* dan Optimasi Naive Bayes menggunakan PSO bernilai 0,941 dengan tingkat diagnosa *Excellent Classification*, didapatkan selisih nilai AUC sebesar 0,047. Sehingga pada penelitian ini, penerapan teknik optimasi Particle Swarm Optimization mampu menyeleksi atribut pada Naive Bayes dan menghasilkan tingkat akurasi diagnosis penyakit hepatitis yang lebih baik dibanding dengan menggunakan metode individual Naive Bayes.

## **2.2 Gambaran Instansi**

Pada subbab ini akan dijelaskan gambaran secara umum dari perusahaan/instansi pemerintah sebagai tempat penelitian tugas akhir. Pokok-pokok yang dijelaskan tentang profil serta pelayanan yang disediakan oleh perusahaan/instansi.

### **2.2.1 Profil Instansi**

Rumah Sakit Ibu & Anak Aisyiyah, atau dikenal sebagai RS Bersalin Aisyiyah yang berada di Jalan Letnan Ramli No 21 Sak Sak, Bangkalan, Jawa Timur merupakan rumah sakit bersalin yang didirikan oleh Aisyiyah, sebuah organisasi perempuan persyarikatan Muhammadiyah. RS Bersalin Aisyiyah Bangkalan didirikan pada tahun 2006 sebagai lembaga pelayanan kesehatan pertama di bidang obstetri dan ginekologi.

### **2.2.2 Jenis Layanan Instansi**

RS Bersalin Aisyiyah membuka pelayanan setiap hari (24 jam), adapun beberapa jenis pelayanan yang diberikan oleh RS Bersalin Aisyiyah, antara lain pemeriksaan kehamilan, meliputi pemeriksaan rutin/bulan dan USG kandungan maupun pemeriksaan kandungan penyakit, pelayanan persalinan normal atau dengan tindakan seperti *sectio caesarea*, *kuretase*, *operasi laporotomi* serta tindakan darurat bagi pasien yang mengalami pendarahan *abortus spontan*. Selain itu juga melayani konsultasi kandungan keluarga berencana (penggunaan alat kontrasepsi), kesulitan punya anak (*infertill*), gangguan haid, gangguan reproduksi (Ediningtias & Candra Premananto, n.d.) .

### **2.3 Kehamilan**

Kehamilan didefinisikan sebagai fertilisasi atau penyatuan dari spermatozoa dan ovum dan dilanjutkan dengan nidasi atau implantasi. Bila dihitung dari saat fertilisasi hingga lahirnya bayi, kehamilan normal akan berlangsung dalam waktu 40 minggu atau 10 bulan atau 9 bulan menurut kalender internasional (Prawirohardjo, 2014). Kehamilan terbagi menjadi 3 trimester, dimana trimester satu berlangsung dalam 12 minggu, trimester kedua 15 minggu dimulai dari minggu ke 13 sampai minggu ke 27, dan trimester ketiga 13 minggu dimulai dari minggu ke 28 sampai minggu ke 40 (Walyani, 2015) .

Kehamilan resiko tinggi merupakan suatu kehamilan yang mempunyai risiko lebih besar dari biasanya, umumnya kehamilan (baik itu bagi ibu maupun bayinya) dimana kehamilan resiko tinggi ini dapat menyebabkan terjadinya bahaya dan

komplikasi selama masa kehamilan, persalinan, ataupun nifas bila dibandingkan dengan kehamilan persalinan dan nifas normal (Maryunani, 2016).

Bahaya yang dapat terjadi pada kelompok resiko ini seperti persalinan berjalan tidak lancar dan bayi sukar lahir (dalam bahaya). Dalam hal ini, pertolongan dapat diberikan oleh PKK dan tenaga kesehatan dengan memberi komunikasi, informasi, edukasi untuk memeriksakan kehamilan secara teratur dan juga membuat perencanaan persalinan dengan ibu hamil, suami dan keluarga untuk melahirkan dengan tepat.

#### **2.4 Persalinan**

Persalinan adalah proses pengeluaran hasil konsepsi (janin dan uri) yang telah cukup bulan atau dapat hidup diluar kandungan melalui jalan lahir atau dengan bantuan atau tanpa bantuan yaitu kekuatan sendiri (Istiqomah & Marmi, 2016). Persalinan adalah proses pengeluaran hasil konsepsi yang dapat hidup dari dalam uterus ke dunia luar. Persalinan dan kelahiran normal merupakan proses pengeluaran janin yang terjadi pada kehamilan cukup bulan (37-42 minggu), lahir spontan dengan presentasi belakang kepala yang berlangsung dalam waktu 18 jam, tanpa komplikasi baik ibu maupun janin (Jannah, 2015). Macam-macam persalinan menurut (Manuaba, 1998) adalah:

1. Persalinan spontan yaitu bila persalinan berlangsung dengan tenaga sendiri.
2. Persalinan buatan yaitu bila persalinan dengan rangsangan sehingga terdapat kekuatan untuk persalinan.
3. Persalinan anjuran yaitu yang paling ideal sudah tentu persalinan spontan karena tidak memerlukan bantuan apapun yang mempunyai trauma

persalinan yang paling ringan sehingga kualitas sumber daya manusia dapat terjamin.

Jenis persalinan yang digunakan dalam penelitian ini yaitu persalinan normal yaitu proses pengeluaran hasil konsepsi (janin) yang dapat hidup dari dalam uterus melalui vagina ke dunia luar secara spontan tanpa bantuan alat dan tidak melukai ibu dan janin yang berlangsung sekitar 18-24 jam, dengan letak janin belakang kepala dan persalinan *sectio caesarea* adalah suatu cara melahirkan janin dengan membuat sayatan pada dinding depan perut atau vagina (Komputer et al., 2020).

Banyak faktor yang dapat mempengaruhi lamanya proses persalinan berlangsung bahkan dapat mempengaruhi keberhasilan melahirkan secara normal, menurut Hebert Hutabarat sebagaimana dikutip oleh (Manuaba, 1998) faktor-faktor tersebut antara lain:

1. Umur kurang dari 19 tahun atau diatas 35 tahun
2. Paritas atau jumlah atau banyaknya persalinan yang pernah dialami ibu baik lahir hidup maupun mati.
3. Riwayat persalinan seperti :
  - Abortus lebih dari 2 kali,
  - Partus prematur 2 kali atau lebih,
  - Riwayat kematian janin dalam rahim,
  - Perdarahan pascapersalinan,
  - Riwayat pre-eklampsia-eklampsia,
  - Riwayat kehamilan mola hidatidosa,
  - Riwayat persalinan dengan tindakan operasi (ekstraksi vakum, ekstraksi forseps, ekstraksi versi, atau plasenta manual),

- Terdapat disproporsi sefalopelvik,
- Perdarahan antepartum,
- Kehamilan ganda atau hidramnion,
- Hamil dengan kelainan letak,
- Sangkaan dimaturitas,
- Serviks inkompeten dan
- Hamil disertai mioma uteri atau kista ovarium.

Faktor-faktor mempengaruhi persalinan yang digunakan dalam penelitian ini adalah usia ibu, usia kehamilan, tinggi badan, riwayat bedah sesar, riwayat obsteri, paritas, tekanan darah, letak sungsang, *cephalopelvic disproportion (CPD)*, *plasenta previa*, preeklamsia berat, oligohidroamnion, jarak kelahiran dan kekuatan ibu (*power*).

#### 1. Usia

Menurut (Sumelung et al., n.d.) kehamilan di usia muda atau remaja di bawah usia 20 tahun akan mengakibatkan rasa takut dan persalinan, hal tersebut dikarenakan ibu belum siap memiliki anak serta alat reproduksi ibu yang belum siap untuk hamil. Pada usia diatas 35 tahun, kematangan organ reproduksi mengalami penurunan sehingga dapat menimbulkan masalah kesehatan pada saat persalinan dan beresiko terjadinya persalinan preterm (Anasari & Pantiawati, 2016). Usia 20-35 tahun adalah usia yang paling tepat bagi wanita untuk memiliki anak, karena pada usia tersebut baik secara psikologis maupun biologis ibu telah siap untuk mengalami kehamilan dan persalinan (Nadhifa Anwar & Tutik, 2018).

#### 2. Usia Kehamilan

Lamanya kehamilan mulai dari ovulasi sampai partus adalah kira – kira 280 hari (40 minggu) dan tidak lebih dari 300 hari (43 minggu), kehamilan 37 - 42 minggu ini disebut kehamilan matur atau cukup bulan (Prawirohardjo, 2014). Sedangkan ketuban pecah dini adalah pecahnya ketuban sebelum waktunya melahirkan, hal ini dapat terjadi pada akhir kehamilan maupun jauh sebelum waktunya melahirkan. Sehingga pecahnya ketuban sebelum usia kehamilan kurang dari 37 minggu disebut preterm. Kehamilan lebih dari 42 minggu disebut kehamilan postmatur/postterm , hal ini akan berkaitan dengan peningkatan kejadian gawat janin dengan risiko tiga kali lebih tinggi. Akibatnya proses penuaan plasenta maka pemasukan makanan dan oksigen menurun sehingga janin akan mengalami pertumbuhan terhambat dan penurunan berat. Jumlah air ketuban yang berkurang mengakibatkan juga perubahan abnormal jantung janin hingga kematian.

### 3. Tinggi badan

Tinggi badan adalah salah satu indikator pertumbuhan. Tinggi badan dipengaruhi oleh beberapa faktor yaitu faktor internal, eksternal dan lingkungan. Perempuan memiliki panggul yang lebih lebar dibandingkan bahu dan kedua spina iliaca anterior superior (SIAS) terpisah dengan jarak yang lebih lebar, sedangkan laki-laki memiliki bahu yang lebih lebar dibandingkan panggul dan kedua SIAS terpisah dengan jarak yang tidak begitu lebar. Tinggi badan ibu dapat memprediksi risiko terhambatnya persalinan yang merupakan penyumbang utama morbiditas, mortalitas ibu dan perinatal di negara berkembang. Perawakan ibu dengan tinggi < 145 cm dikaitkan dengan hasil kehamilan yang merugikan, seperti bayi lahir mati, bayi lahir dengan berat badan

lahir rendah, bahkan kematian perinatal. Meskipun tinggi badan ibu dapat memprediksi risiko persalinan yang terhambat, hal itu juga tergantung dari indeks kesehatan umum dan status gizi wanita dari masa kanak-kanaknya, di mana faktor genetik berperan menyokongnya. Dengan demikian, signifikansi obstetrik dari tinggi badan ibu dikaitkan dengan latar belakang genetik pasien sendiri (Humaera et al., 2018).

#### 4. Riwayat bedah sesar

Riwayat SC merupakan ibu hamil dengan persalinan yang sebelumnya dilakukan operasi sesar. Oleh karena itu pada dinding rahim ibu terdapat cacat bekas luka operasi. Bahaya pada robekan rahim yaitu kematian janin dan kematian ibu, perdarahan dan infeksi (Widatiningsih & Dewi, 2017). Persalinan secara sesar pada ibu dengan riwayat sesar memiliki resiko kematian lebih tinggi tetapi persalinan normal pada ibu dengan riwayat sesar memiliki resiko komplikasi yang lebih tinggi, komplikasi yang dapat dialami ialah perdarahan dan rupture (Candy et al., 2018).

#### 5. Riwayat Obsteri

Pemeriksaan obstetri meliputi banyak prosedur yang masing-masing berkaitan dengan tujuan pemeriksaan yang dilakukan. Untuk pemeriksaan dasar obstetri, pada umumnya diperlukan pemeriksaan antenatal, pemeriksaan fisik ibu hamil meliputi inpeksi, palpasi dan auskultasi. Pemeriksaan sederhana ini dapat mendeteksi dini berbagai kondisi patologis yang dapat menyulitkan kehamilan. Riwayat Obsteri Buruk (ROB) dapat terjadi pada Ibu hamil dengan kehamilan kedua, dimana kehamilan yang pertama mengalami keguguran, lahir belum cukup bulan, lahir mati, lahir hidup lalu mati umur < 7 hari. Kehamilan ketiga

atau lebih, kehamilan yang lalu pernah mengalami keguguran >2 kali. Kehamilan kedua atau lebih, kehamilan terakhir janin mati dalam kandungan.

#### 6. Paritas

Paritas adalah jumlah atau banyaknya persalinan yang pernah dialami ibu baik lahir hidup maupun mati. Ibu dengan paritas pertama akan berisiko karena rahim baru pertama kali menerima hasil konsepsi dan keluwesan otot rahim masih terbatas untuk pertumbuhan janin sedangkan ibu dengan paritas tinggi atau lebih dari 3 memiliki angka maternal yang tinggi karena dapat terjadi gangguan endometrium. Penyebab gangguan endometrium tersebut dikarenakan kehamilan berulang. Sedangkan paritas dapat dibedakan menjadi multipara yaitu paritas 0, primipara yaitu paritas 1, multipara yaitu paritas 2-4, dan grandemultipara yaitu paritas lebih dari 4 (Prawirohardjo, 2014).

#### 7. Tekanan darah

Tekanan darah normal yaitu 120/80 mmHg. Bila tekanan darah lebih besar atau sama dengan 140/90 mmHg, ada faktor risiko hipertensi (tekanan darah tinggi) dalam kehamilan.

#### 8. Letak sungsang

Letak sungsang merupakan keadaan dimana janin terletak memanjang dengan kepala di fundus uteri dan bokong berada di bagian bawah kavum uteri (Aprina & Puri, 2016). Idealnya, bayi lahir dengan kepala terlebih dahulu. Akan tetapi, terkadang bayi justru lahir dengan bokong terlebih dahulu, kemudian kepala. Posisi bayi sungsang dapat menyebabkan permasalahan saat akan menjalani persalinan.

#### 9. Cephalopelvic Disproportion (CPD)

Cephalopelvic Disproportion atau pinggul sempit adalah disproporsi antara ukuran janin dan ukuran pelvis, yakni ukuran pelvis tertentu tidak cukup besar untuk mengakomodasi keluarnya janin tertentu melalui pelvis sampai terjadi kelahiran pervaginam (Varney et al., 2007).

#### 10. Plasenta Previa

Plasenta Previa adalah keadaan dimana plasenta berimplantasi pada tempat abnormal, yaitu pada segmen bawah Rahim sehingga menutupi sebagian atau seluruh pembukaan jalan lahir. Pada keadaan normal plasenta terletak di bagian atas uterus (Megasari, 2015).

#### 11. Preeklamsia Berat

Preeklamsia adalah penyakit dengan tanda-tanda hipertensi, edema dan proteinuria yang timbul karena kehamilan. Preeklamsia dibagi menjadi 2 golongan yaitu ringan dan berat. Preeklamsia dapat digolongkan menjadi preeklamsia berat apabila terdapat satu atau lebih gejala di bawah ini ditemukan (Permatasari, 2019):

- Tekanan sistolik 160 mmHg, atau tekanan diastolik 110 mmHg
- Proteinuria 3 atau 4+ pada pemeriksaan kualitatif
- Oliguria (air kencing) 400mL atau kurang dalam 24 jam
- Keluhan serebral, gangguan penglihatan atau nyeri di daerah epigastrium
- Edema paru atau sianosis

#### 12. Oligohidroamnion

Kondisi volume air ketuban sangat sedikit. Hal ini disebabkan karena janin sedikit sekali dalam memproduksi air ketuban. Selain itu juga dapat disebabkan

air ketuban yang merembes keluar. Air ketuban yang merembes ini biasanya karena selaput ketuban yang membungkusnya robek (Permatasari, 2019).

### 13. Jarak Kelahiran

Ibu hamil yang jarak kelahiran dengan anak terkecil kurang dari 2 tahun. Kesehatan fisik dan Rahim ibu masih butuh cukup istirahat. Ada kemungkinan ibu masih menyusui. Anak masih butuh asuhan dan perhatian orang tuanya.

### 14. Kekuatan Ibu (*Power*)

Induksi persalinan yaitu suatu tindakan yang dilakukan terhadap ibu hamil yang belum dalam persalinan untuk merangsang terjadinya persalinan. Faktor yang bisa diinisiasi agar induksi persalinan dapat berhasil adalah matangnya serviks. Penilaian kematangan serviks dengan menggunakan *Bishop Score*. Hasil penilaian akan berpengaruh pada keberhasilan induksi persalinan. Hasil Bishop Score kurang dari 5 risiko terjadi induksi gagal. Induksi gagal diartikan sebagai kegagalan timbulnya persalinan dalam satu siklus terapi, solusi pada kasus kegagalan induksi adalah dengan meneruskan induksi atau melakukan persalinan Sectio Caesarea (Retnaningsih et al., 2018).

## 2.5 *Data Mining*

*Data mining* merupakan proses untuk menggali (*mining*) pengetahuan dan informasi baru dari data yang berjumlah banyak pada data warehouse, dengan menggunakan kecerdasan buatan (*Artificial Intelligence*), statistik dan matematika. *Data mining* merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pemakainya (Jollyta et al., 2020).

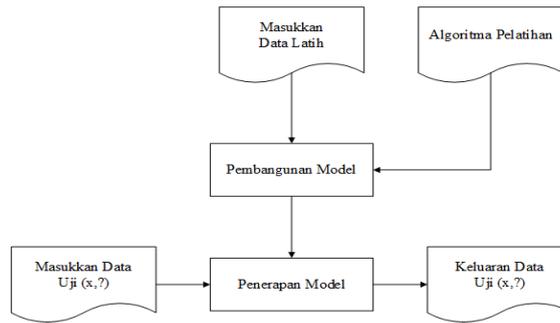
Proses pengolahan data dalam data mining memerlukan sebuah algoritma untuk mengekstraksi data menjadi sebuah informasi yang tiap-tiap algoritma

dikelompokkan berdasarkan masing-masing peranan, pada kasus estimasi dan prediksi banyak digunakan algoritma Linear Regression, Support Vector Machine, Neural Network, sedangkan untuk klasifikasi banyak yang menggunakan algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree (ID3, C4.5, C5.0, CART). Dan dalam kasus clustering, dapat digunakan algoritma K-Means, Fuzzy, C-Means, K-Medoid, Self-Organization Map. Untuk asosiasi digunakan algoritma FP-Growth, A priori, Chi Square, dan Coefficient of Correlation (Suntoro, 2019).

## **2.6 Konsep Klasifikasi**

Klasifikasi merupakan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototype untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya (Prabowo et al., 2018).

Klasifikasi adalah metode data mining yang dapat digunakan untuk proses pencarian sekumpulan model (fungsi) yang dapat menjelaskan dan membedakan kelas-kelas data atau konsep, yang tujuannya supaya model tersebut dapat digunakan memprediksi objek kelas yang labelnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang muncul di masa depan. Proses klasifikasi tersebut seperti terlihat pada Gambar 2.1.



**Gambar 2. 1** Proses Pekerjaan Klasifikasi

## 2.7 Particle Swarm Optimazation

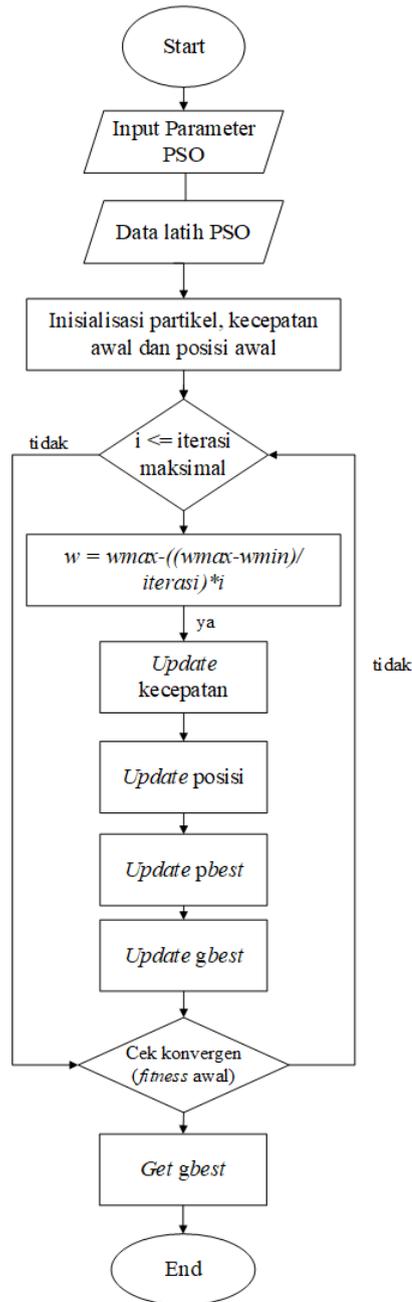
Particle Swarm Optimization atau PSO ditemukan oleh Russell C. Eberhart dan James Kennedy di tahun 1995, merupakan salah satu algoritma yang ada pada *swarm intelegence*. Metode ini meniru cara kerja sekelompok burung. Pembelajaran partikel terdiri dari *cognitive learning* atau *pBest* (posisi terbaik yang pernah dicapai partikel) dan *social learning* atau *gBest* (posisi terbaik dari keseluruhan partikel dalam kelompok) (Novitasari et al., 2016). Pada algoritma *Particle Swarm Optimization*, pencarian solusi dilakukan oleh suatu populasi yang terdiri dari beberapa partikel, tiap partikel berpindah dengan kecepatan yang diadaptasi dari daerah pencarian dan menyimpannya sebagai posisi terbaik yang pernah dicapai, dalam PSO populasi disebut dengan *swarm*, dan individu disebut dengan *particle* (Mutiara, 2020).

Kata partikel menunjukkan, misalnya, seekor burung dalam kawanan burung dan setiap individu atau partikel berperilaku secara terdistribusi dengan cara menggunakan kecerdasannya (*intelligence*) sendiri dan juga dipengaruhi perilaku kelompok kolektifnya. 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 di kelompok tersebut. Dalam konteks optimasi multivariabel,

kawanan diasumsikan mempunyai ukuran tertentu atau tetap dengan setiap partikel posisi awalnya terletak di suatu lokasi yang acak dalam ruang multidimensi. Setiap partikel bergerak dalam ruang/space tertentu dan mengingat posisi terbaik yang pernah dilalui atau ditemukan terhadap sumber makanan atau nilai fungsi objektif. Setiap partikel menyampaikan informasi atau posisi bagusnya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang bagus tersebut, misalnya perilaku burung-burung dalam kawanan burung. Meskipun setiap burung mempunyai keterbatasan dalam hal kecerdasan, biasanya dia akan mengikuti kebiasaan (rule) seperti berikut:

1. Seekor burung tidak berada terlalu dekat dengan burung yang lain
2. Burung tersebut akan mengarahkan terbangnya ke arah rata-rata keseluruhan burung
3. Akan memposisikan diri dengan rata-rata posisi burung yang lain dengan menjaga, sehingga jarak antar burung dalam kawanan itu tidak terlalu jauh.

Model ini akan disimulasikan dalam ruang dengan dimensi tertentu dengan sejumlah iterasi sehingga di setiap iterasi, posisi partikel akan semakin mengarah ke target yang dituju (minimum atau maksimum fungsi). Ini dilakukan hingga iterasi telah mencapai iterasi maksimum atau PSO telah menemukan nilai optimum tertentu atau kesalahan minimum yang diinginkan. Supaya lebih jelas, berikut diagram alir proses dari algoritma PSO .



**Gambar 2. 2** Flowchart PSO

Algoritma dasar PSO terdiri dari tiga tahap, yaitu pembangkitan posisi serta kecepatan partikel, *update velocity* (update kecepatan), *update position* (update posisi). Partikel berubah posisinya dari suatu perpindahan (iterasi) ke posisi lainnya berdasarkan pada *update velocity*. Pertama posisi  $x_0^i$ , dan kecepatan  $v_k^i$  dari kumpulan partikel dibangkitkan secara random menggunakan batas atas ( $x_{max}$ )

dan batas bawah ( $x_{min}$ ) dari *design variable*, seperti yang ditunjukkan pada persamaan (2.4) dan (2.5) .

$$x_0^i = x_{min} + rand (x_{max} - x_{min}) \quad (2.1)$$

$$v_0^i = x_{min} + rand (x_{max} - x_{min}) \quad (2.2)$$

Keterangan:

$x_0^i$  = posisi awal

$v_0^i$  = kecepatan awal

$x_{min}$  = batas bawah

$x_{max}$  = batas atas

*rand* = nilai *random* antara nilai 0 dan 1

Posisi dan kecepatan direpresentasikan dalam bentuk vektor dimana  $n$  dimensi vektor merepresentasikan jumlah dari desain variabel partikel, dengan superscript dan subscript menotasikan partikel ke  $i$  pada waktu ke  $k$ . Dengan proses inisialisasi ini maka kumpulan partikel dapat terdistribusi secara random pada *design space*. Vektor seperti ditunjukkan dibawah ini :

$$x_k^i = (x_k^{i1}, x_k^{i2}, \dots \dots \dots, x_k^{in})^T \quad (2.3)$$

$$v_k^i = (v_k^{i1}, v_k^{i2}, \dots \dots \dots, v_k^{in})^T \quad (2.4)$$

Langkah kedua adalah *update velocity* (kecepatan) untuk semua partikel pada waktu  $k + 1$  menggunakan fungsi objektif atau nilai fitness posisi partikel saat ini pada *design space* saat waktu ke  $k$ . Dari nilai fitness dapat ditentukan partikel mana yang memiliki nilai global terbaik (*global best*) pada *swarm* saat ini,  $P_k^g$ , dan juga dapat ditentukan posisi terbaik dari tiap partikel pada semua waktu yang sekarang dan sebelumnya,  $P^i$ . Perumusan *update velocity* menggunakan dua informasi tersebut untuk semua partikel pada kumpulan dengan pengaruh perpindahan yang sekarang,  $v_k^i$ , untuk memberikan arah pencarian,  $v_{k+1}^i$ , untuk generasi selanjutnya.

Perumusan *update velocity* mencakup beberapa parameter random, *rnd* , untuk mendapatkan cakupan yang baik pada *design space*, tiga parameter yang mempengaruhi arah pencarian, yaitu *inertia factor* (*w*), *self confidence* (*c1*), *swarm confidence* (*c2*) akan digabungkan dalam satu penyajian, seperti yang ditunjukkan persamaan berikut:

$$v_{k+1}^i = (w * v_k^i + c1 * rnd * (p^i - x_k^i)) + c2 * rnd * (p_k^g - x_k^i) \quad (2.5)$$

Keterangan :

*w* = mengontrol pengaruh kecepatan sebelumnya dikecepatan sekarang dengan range  $w = 0.4 - 1.4$

*c1,c2* = *learning rates* untuk kemampuan individu (*cognitive*) dan pengaruh sosial (*group*) dengan  $c1 = 0.5 - 2.0$ ,  $c2 = 0.5 - 2.0$  .

$p^i$  = *local best*, posisi terbaik dari tiap partikel pada semua waktu yang sekarang

$p_k^g$  = nilai global terbaik (*global best*) pada *swarm* saat ini

$v_k^i$  = kecepatan sekarang

$x_k^i$  = posisi sekarang

Langkah terakhir dari setiap iterasi adalah *update posisi* tiap partikel dengan *vektor velocity*, seperti yang ditunjukkan pada persamaan (2.6) berikut ini :

$$x_{k+1}^i = x_k^i + v_{k+1}^i \quad (2.6)$$

Ket :

$x_{k+1}^i$  = posisi pencarian

$x_k^i$  = posisi sekarang

$v_{k+1}^i$  = arah pencarian

Tiga tahapan diatas akan diulang sampai kriteria kekonvergenan terpenuhi, kriteria kekonvergenan sangat penting dalam menghindari penambahan fungsi evaluasi setelah solusi optimum didapatkan, namun kriteria kekonvergenan tidak

selalu mutlak diperlukan, penetapan jumlah iterasi maksimal juga dapat digunakan sebagai *stopping condition* dari algoritma.

### 2.7.1 Fungsi Fitness pada PSO

Fungsi *fitness* digunakan untuk mengukur tingkat kebaikan atau kesesuaian (*fitness*) suatu solusi dengan solusi yang dicari. Fungsi *fitness* bisa berhubungan langsung dengan fungsi tujuan, atau bisa juga sedikit modifikasi terhadap fungsi tujuan. Sejumlah solusi yang dibangkitkan dalam populasi akan dievaluasi menggunakan fungsi *fitness*.

Untuk menghitung nilai fitness, terlebih dahulu menghitung *cost*. *Cost* atau penalti adalah total ketidaksesuaian antara kelas data latih PSO hasil klasifikasi terhadap kelas data latih target atau kelas sesungguhnya (Setyowati & Mahmudy, 2018). Untuk menghitung nilai *cost* dapat dilakukan dengan cara sebagai berikut :

1. Menentukan data latih
2. Untuk setiap partikel lakukan langkah 3 sampai 4
3. Untuk setiap data training lakukan langka a sampai c
  - a. Hitung jarak euclidean antara data latih terhadap dimensi 0-13 partikel menggunakan persamaan :

$$D(x,y) = ||x_i - y_i|| \quad (2.7)$$

- b. Hitung jarak euclidean antara data latih terhadap dimensi 14-27 partikel menggunakan persamaan :

$$D(x,y) = ||x_i - y_i|| \quad (2.8)$$

- c. Cari kelas kalsifikasi dengan cara mencari euclidean terkecil antara langka a sampai b dengan persamaan :

$$E = \text{Min } D(x,y) \quad (2.9)$$

d. Perbarui nilai *cost* dengan syarat :

Jika  $T = C$

$$\text{Cost} = \text{cost} + 1 \quad (2.10)$$

Jika  $T \neq C$

$$\text{Cost} = \text{cost} + 0 \quad (2.11)$$

4. Hitung nilai *fitness* partikel dengan persamaan :

$$\text{fitness} = \frac{\text{data latih PSO} - \text{cost}}{\text{banyak data latih PSO}} \quad (2.12)$$

5. Berhenti jika semua partikel sudah mendapatkan *fitness*

## 2.8 Naïve Bayes

Metode Naive Bayes merupakan salah satu metode klasifikasi yang efektif dan efisien karena proses pengklasifikasiannya bekerja secara independen pada setiap fitur objek yang akan diklasifikasi (Aji, 2019). Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif), dengan kata lain yakni Naïve Bayes model yang digunakan adalah “model fitur independen” .

Metode Naive Bayes Classifier menggunakan konsep probabilitas yang bertujuan untuk melakukan klasifikasi data pada class tertentu, metode Naive Bayes Classifier merupakan penyederhanaan dari teorema bayes. Prediksi bayes didasarkan pada teorema bayes dengan formula umum pada persamaan berikut (Satya & Hidayat, 2018) :

1. Menentukan nilai prior untuk tiap kelas dengan membandingkan banyak anggota suatu kelas dengan keseluruhan data sampel yaitu menggunakan persamaan berikut :

$$P(H) = \frac{X}{A} \quad (2.13)$$

Ket :  $P$  = Nilai Prior

$X$  = Jumlah data tiap kelas

$A$  = Jumlah data seluruh kelas

2. Menentukan nilai likelihood dengan menentukan nilai probabilitas setiap atribut terhadap kelasnya, kemungkinan kemunculan kelas ketika suatu atribut terpilih yaitu menggunakan persamaan sebagai berikut :

$$P(E/H) = \frac{F}{B} \quad (2.14)$$

Ket :  $P(E/H)$  = Nilai Likelihood

$F$  = Jumlah data fitur tiap kelas

$B$  = Jumlah seluruh data tiap kelas

3. Menentukan posterior yaitu sebuah hasil perhitungan likelihood dalam bentuk probabilitas atribut terhadap kelas digunakan untuk mencari peluang masuknya sampel karakteristik tertentu dalam suatu kelas. Perhitungan posterior dihitung dengan cara mengalikan kemungkinan atribut masukan dengan kelas, pada proses ini didapat probabilitas akhir untuk kesimpulan hasil persalinan :

$$P(H/E) = P(H) \times P(E/H) \quad (2.15)$$

Ket :  $P(H/E)$  = Nilai Posterior

$P(H)$  = Nilai Prior

$$P(E/H) = \text{Nilai likelihood}$$

Hasil klasifikasi dilakukan dengan cara membandingkan nilai posterior dari kelas kelas yang ada. Nilai posterior yang paling tinggi adalah yang terpilih menjadi hasil klasifikasi.

## 2.9 Confusion Matrix & Receiver Operating Characteristic (ROC) Curve

*Confusion matrix* merupakan sebuah metode untuk mengukur atau evaluasi atas model yang telah dibuat dengan menghitung nilai *performance metrics*, diantaranya *accuracy*, *precision*, dan *recall*. Akurasi menyatakan jumlah data yang diklasifikasikan benar setelah dilakukan proses pengujian, sedangkan laju error digunakan untuk menghitung kesalahan identifikasi (Kastawan, Wiharta, & Sudarma, 2018).

*Confusion matrix* biasa digunakan pada algoritma klasifikasi (*supervised learning*) *binary class* maupun *multi class* yang pada dasarnya memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh model prediksi dengan hasil klasifikasi yang sebenarnya. Setiap kolom pada matriks adalah hasil dari kelas prediksi, sedangkan tiap baris berisi kejadian di kelas yang sebenarnya sehingga *confusion matrix* terdiri dari informasi aktual dan prediksi pada sistem klasifikasi, seperti pada tabel 2.1 .

**Tabel 2. 1** Tabel Confusion Matrix

	Kelas	Prediksi	
		A	B
Aktual	A	TP	FP
	B	FN	TN

Dari tabel *confusion matrix* diatas, terdapat 4 jenis nilai yaitu TP (*True Positive*), FP (*False Positive*), FN (*False Negative*), dan TN (*True Positive*) yang

nantinya akan menghasilkan 3 nilai, diantaranya nilai *recall*, *precision*, dan *accuracy*. *Recall* merupakan rasio kasus positif yang diidentifikasi dengan benar sedangkan *precision* adalah rasio kasus yang diprediksi benar positif dibanding dengan keseluruhan data prediksi positif serta *accuracy* merupakan rasio prediksi benar dengan keseluruhan data. Untuk mendapatkan ketiga nilai tersebut dapat dengan menggunakan persamaan dibawah ini.

$$recall = \frac{TP}{TP+FN} * 100\% \quad (2.16)$$

$$precision = \frac{TP}{TP+FP} * 100\% \quad (2.17)$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (2.18)$$

Sedangkan fungsi ROC Curve adalah untuk memperlihatkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan Confusion Matrix. ROC merupakan grafik dua dimensi dengan garis horizontal sebagai *false positive* dan garis vertical sebagai *true positive*. Secara teknis kurva ROC juga dikenal sebagai grafik ROC, dua dimensi grafik dimana tingkat TP diplot pada sumbu Y dan tingkat FP diplot pada sumbu X. Hasil perhitungan dapat divisualisasikan dengan ROC Curve atau AUC (*Area Under Curve*). Berikut tingkat nilai diagnosa dari ROC Curve :

1. Akurasi bernilai 0.90-1.00 sama dengan Excellent Classification
2. Akurasi bernilai 0.80-0.90 sama dengan Good Classification
3. Akurasi bernilai 0.70-0.80 sama dengan Fair Classification
4. Akurasi bernilai 0.60-0.70 sama dengan Poor Classification
5. Akurasi bernilai 0.50-0.60 sama dengan Failur