

BAB II

LANDASAN TEORI

2.1 Penelitian Terkait

Beberapa penelitian terkait dengan menggunakan teknik Data Mining dengan penerapan beberapa algoritma yang berkaitan dengan karyawan diantaranya:

Tabel 2. 1 Review Jurnal

No	Judul, Penulis, Tahun	Dataset	Metode	Hasil	Kekurangan	Kelebihan
1	Penerapan klasifikasi status pegawai menggunakan metode naïve bayes di RSUD H. Syaiful Anwar, Rino Gupitha, 2018	Data set pada penelitian ini diambil dari RSUD H. Syaiful Anwar, berupa data kuantitatif	Menggunakan Metode <i>Naive Bayes</i>	Dari 7 atribut yang terdapat dalam dataset yaitu nama, lama kerja, level jabatan, level pendidikan, pengalaman, gaji pokok, status pegawai. Penelitian ini menggunakan Algoritma <i>Naive Bayes</i> .	1. Tidak ada grafik 2. Dalam penelitian, ada terlalu banyak konten teoretis dalam diskusi 3. Hasil akurasi masih belum optimal yaitu <i>Naive Bayes</i> sebesar 72%	- Jurnal menggunakan bahasa yang mudah dipahami

2	Penentuan Jadwal Kerja Berdasarkan Klasifikasi Data Karyawan Menggunakan Metode Decision Tree Algorithm C4.5, Mas'ud Effendi, 2021	Company data, observation and documentation	DECISION TREE C4.5	Pengujian sistem menggunakan K-fold cross validation menghasilkan rata-rata akurasi sebesar 93.39%	1. Memberikan info detail dataset 2. Tidak ada grafik.	1. Pembahasan jurnal cukup lengkap. 2. Jurnal menggunakan bahasa yang mudah dipahami.
3	Prediksi Pengunduran Diri Karyawan Perusahaan "Y" Menggunakan Random Forest, Daniel Dwif Eryanto Manurung, 2021	Data set pada penelitian ini berupa data kuantitatif	Random Forest	Pada penelitian ini menggunakan metode Random Forest, hasil akurasi yang didapatkan sebesar 87% dan error sebesar 13%.	1. Dataset kurang jelas 2. Hasil akurasi kurang optimal. 3. Tidak dicantumkan atribut yang terlibat pada dataset	1. Jurnal menggunakan data kuantitatif yang mudah dipahami 2. Jurnal menggunakan bahasa yang mudah dipahami 3. terdapat grafik yang mudah untuk dibaca dan dipahami.
4	Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5, Ikhsan Romli, 2020	Data set pada penelitian ini berupa data kuantitatif	Algoritma C4.5	Dalam penelitian ini menggunakan algoritma C4.5 untuk menghasilkan jadwal overtime. Hasil uji coba jadwal overtime dengan algoritma C4.5 dengan Confusion matrix memiliki nilai	1. Sumber dataset tidak diketahui 2. Perhitungan manual yang ditampilkan terlalu susah dimengerti	1. Penulis menampilkan atribut yang digunakan 2. Hasil pohon keputusan ada sehingga pembaca bisa melihat sebagai pengambil keputusan.

				accuracy, precision, dan recall yang bagus yaitu dengan accuracy 91%, precision 86.05% dan recall 92.5%.		3. Jurnal menggunakan bahasa yang mudah dipahami.
5	An Analysis of Contract Employee Performance Assessment Using Machine Learning, Mangapul Siahaan, 2021	PT. XYZ merupakan perusahaan manufaktur di Kota Batam yang mempekerjakan karyawan kontrak atau PKWT (perjanjian kerja waktu tertentu)	Naives Bayes, Logistik Regression, Decision Tree, K-Nearest Neighbors, Random Forest	Berdasarkan penelitian, model algoritma Random Forest yang terbaik digunakan dalam mengklasifikasi dan prediksi assement kinerja karyawan kontrak dengan mencapai nilai akurasi 90.62 %, presisi 72.22 %, recall 75% dan nilai F1 score 71.11 %. Hasil : Decision Tree akurasi 84.38%, Naïvet Bayes 78.12 %, Random Forest 90.62 %, K-Nearest Neighbors 56.25 %, Logistik Regression 62.5 %.	Perbandingan algoritma pada jurnal ini menentukan mana algoritma yang terbaik untuk digunakan diambil dari akurasi yang paling besar. Kesimpulan dari penelitian ini adalah algoritma random forest yang memiliki akurasi tertinggi disusul oleh algoritma decision tree.	1. Penulis menampilkan setiap hasil perbandingan algoritma 2. menggunakan bahasa yang mudah dipahami. 3. Terdapat grafik yang mudah dipahami 4. Atribut dan dataset yang diolah cukup rapi dan mudah dipahami.
7	Prediksi Loyalitas dalam Keterikatan Karyawan terhadap Perusahaan Menggunakan Algoritma C4.5*	Studi Kasus PT.XYZ	Algoritma C4.5	Hasil perhitungan confusion matrix dan 10 Ten-Fold Cross Validation dihasilkan Accuracy yang baik sebesar 90,11%	1. Tidak ada grafik 2. Dalam penelitian, ada terlampir banyak konten teoretis dalam diskusi	Penjelasan mudah dipahami, terdapat perhitungan dan gambar yang mendukung proses penelitian.

	(Studi Kasus PT.XYZ)				3. Hasil akurasi masih belum optimal.	
6	Employee Future Prediction, Saim Chandan P Reddy, 2021	Kaggle.com (https://www.kaggle.com/datasets/tejas/hvi14/employee-future-prediction)	Decision Tree	Penelitian ini menggunakan kombinasi pohon keputusan dan regresi linier untuk mencapai akurasi 86,39 persen, mengungguli Adaboost dan Random Forest, as serta SVM.	Penjelasan terlalu singkat, tidak ada perhitungan manual. Tidak ada perbandingan hasil pada setiap algoritma. Hasil kurang maksimal	1. Bahasa mudah dimengerti. 2. Terdapat visualisasi yang lengkap pada setiap atribut 3. Atribut dan dataset jelas.

Dari hasil review beberapa jurnal diatas baik jurnal nasional dan internasional maka dapat disimpulkan untuk mendapatkan akurasi yang baik maka data yang dibutuhkan pada penelitian yaitu minimal 500 record data, dan metode yang digunakan untuk beberapa jurnal yang sudah di review tingkat akurasinya cenderung lebih tinggi menggunakan algoritma *Decision Tree* C4.5 dengan mencapai akurasi 91%. hasil perbandingan akurasi dapat dilihat dibawah ini:

Tabel 2. 2 Hasil Akurasi Penelitian Sebelumnya

No	Jurnal	Algoritma	Akurasi
1	Rino Gupitha 2018	<i>Naive Bayes</i>	72.00 %
2	Mas'ud Effendi, 2021	C4.5	93.39%
3	Daniel Dwi Eryanto Manurung, 2021	<i>Random Forest</i>	87,00%
4	Ikhsan Romli, 2020	C4.5	91.00%
5	Mangapul Siahaan, 2021	Random Forest	90.62 %
6	Sai Chandan P Reddy, 2021	Decision Tree	86,39 %

2.2 Landasan Teori

2.2.1 Karyawan

a. Pengertian Karyawan

Karyawan merupakan aset perusahaan. Kehadiran karyawan begitu sangat penting hingga saat ini, tanpa adanya karyawan tidak akan terjadi kelancaran dan proses produksi suatu perusahaan. Menurut Undang-Undang nomor 13 tahun 2003 tentang ketenaga kerjaan pasal 1 ayat 2 menyebutkan bahwa karyawan adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang dan jasa baik untuk memenuhi kebutuhan sendiri maupun masyarakat, baik di dalam maupun di luar hubungan kerja. Dari definisi tersebut maka yang dimaksud tenaga kerja adalah tenaga kerja yang melakukan pekerjaan pada setiap bentuk usaha (perusahaan) atau perorangan dengan menerima upah termasuk tenaga kerja yang melakukan pekerjaan di luar hubungan kerja. Karyawan merupakan kekayaan utama dalam suatu perusahaan, karena tanpa adanya keikutsertaan mereka, aktifitas tidak akan terlaksana. Karyawan berperan aktif dalam menetapkan rencana, system, proses dan tujuan yang ingin dicapai.

Menurut [8], karyawan merupakan setiap penduduk yang masuk ke dalam usia kerja (15 hingga 64 tahun) atau jumlah total seluruh penduduk yang ada pada sebuah negara yang memproduksi barang dan jasa jika ada permintaan akan tenaga yang akan mereka produksi sendiri dan jika mereka mau berkecimpung/berpartisipasi dalam aktivitas itu.

b. Loyalitas Karyawan

Di banyak karyawan perusahaan, loyalitas telah menjadi hal yang penting. Itu tanggung

jawab SDM (Sumber Daya Manusia) Manajer untuk memenuhi kebutuhan yang diperlukan dari karyawan mereka. Mempekerjakan umumnya meninggalkan perusahaan dan untuk memenuhi tertentu posisi mereka harus menugaskan orang lain untuk tertentu posisi, yang tidak selalu memungkinkan bagi mereka. Sejak kualitas dan keterampilan pekerja mewakili pertumbuhan perusahaan dan keunggulan kompetitif [2].

2.2.2 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Turban, 2005), Proses Data mining menerapkan teknik statistik, machine learning, matematika, dan kecerdasan buatan dalam mengekstraksi dan identifikasi pengetahuan dari berbagai basis data besar. Data mining dapat pula didefinisikan rangkaian proses penggalian nilai tambah dari suatu perkumpulan data yakni nilai informasi yang belum diketahui cara manual.

Definisi umum dari data mining itu sendiri adalah proses pencarian polapola yang tersembunyi (hidden patern) berupa pengetahuan (knowledge) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data werehouse, atau media penyimpanan informasi yang lain. Hal penting yang terkait di dalam data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.

2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat [9].

Data mining memiliki beberapa teknik yang terkenal dan sering digunakan oleh peneliti, diantaranya seperti clustering, classification, association, dan beberapa perkembangan teknik sesuai dengan perubahan kecendrungan data pada saat ini.

Pola yang disajikan mudah dipahami berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, penggalian datanya dengan memiliki beberapa nama alternatif meskipun eksaknya berbeda seperti KDD (*Knowledge Discovery in Database*) , analisis pola, arkeologi data, pemanenan informasi, intelegensia bisnis. Data mining dikelompokkan menjadi beberapa kelompok yaitu:

1. Deskripsi Menggambarkan pola yang terdapat dalam data yang memberikan kemungkinan penjelasan untuk suatu pola atau kecendrungan.
2. Estimasi Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Dibangun dengan record lengkap nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.
3. Prediksi Prediksi nilai dari hasil akan ada di masa mendatang .
4. Klasifikasi Dalam klasifikasi, terdapat target variabel kategori.

5. Pengklusteran Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

2.2.3 Klasifikasi

Klasifikasi adalah urutan yang sangat penting dalam data komunitas pertambangan. Klasifikasi adalah salah satu prediksi teknik data mining yang membuat prediksi tentang data nilai menggunakan hasil yang diketahui yang ditemukan dari kumpulan data yang berbeda. Masalah akurasi dari banyak algoritma klasifikasi adalah diketahui mengalami penurunan informasi saat dihadapi dengan data yang tidak seimbang, misalnya ketika distribusi sampel lintas kelas sangat miring [10]. Dalam klasifikasi, ada variabel kategoris target, seperti braket pendapatan, yang, misalnya, dapat dipartisi menjaditiga kelas atau kategori: berpenghasilan tinggi, menengah pendapatan, dan pendapatan rendah. Model data mining memeriksa satu set besar catatan, masing-masing catatan yang berisi informasi tentangvariabel target serta satu set input atau prediktor variable. Contoh tugas klasifikasi dalam bisnisdan penelitian meliputi : [11].

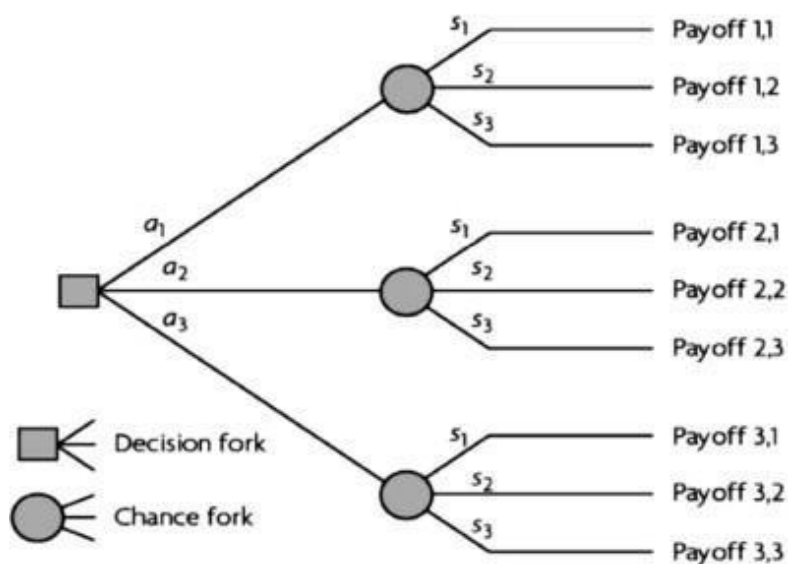
- a. Menentukan apakah transaksi kartu kredit tertentu adalah penipuan
- b. Menempatkan siswa baru pada jalur tertentu yang berkaitan dengan kebutuhan khusus
- c. Menilai apakah aplikasi hipotek adalah risiko kredit yang baik atau buruk
- d. Mendiagnosis apakah ada penyakit tertentu
- e. Menentukan apakah surat wasiat ditulis oleh almarhum yang sebenarnya, atau curangoleh orang lain

- f. Mengidentifikasi apakah perilaku keuangan atau pribadi tertentu menunjukkan kemungkinan ancaman teroris

Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy dan Jaringan Saraf Tiruan [12].

2.2.4 Decision Tree

Decision Tree adalah struktur flowchart yang menyerupai Tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada *Decision Tree* di telusuri dari simpul akar ke simpul daun yang memegang prediksi [13].



Gambar 2. 1 Bentuk *Decision Tree* Secara umum

Decision tree memiliki training sample berupa sekumpulan data yang nantinya akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Secara umum *Decision Tree* adalah untuk membangun pohon keputusan sebagai berikut :

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk setiap nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang yang memiliki kelas yang sama. Rumus menghitung nilai entropy menggunakan persamaan:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \log_2 p_i \quad (2.1)$$

Keterangan :

S = himpunan kasus

n = jumlah partisi atribut A

P_i = proporsi S_i terhadap S

|S_i| = jumlah kasus pada partisi ke i

|S| = jumlah kasus dalam S

A = atribut Rumus untuk mencari nilai gain :

$$\text{Gain (S,A)} = \sum_{i=1}^n \frac{|s_i|}{|s|} \text{Entropy (S}_i) \quad (2.2)$$

2.2.5 *Split Validation*

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian lainnya sebagai data testing. Dengan menggunakan Split Data akan dilakukan percobaan training berdasarkan split ratio yang telah ditentukan sebelumnya, untuk kemudian sisa dari split ratio data training akan dianggap sebagai data testing. Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran [14].

2.2.6 Seleksi Fitur

Seleksi fitur merupakan proses yang melibatkan subset dari kumpulan fitur yang menghasilkan keluaran seperti keseluruhan kumpulan fitur. Seleksi fitur biasanya digunakan untuk memilih fitur yang optimal, mereduksi dimensi, meningkatkan akurasi algoritma klasifier, dan menghapus fitur yang tidak relevan [15]. Tujuan utama dari seleksi fitur adalah untuk mengurangi jumlah fitur yang digunakan dalam klasifikasi dengan tetap menjaga akurasi klasifikasi yang dapat diterima. Pemilihan fitur dapat berdampak besar pada keefektifan algoritma klasifikasi yang dihasilkan, dalam beberapa kasus, sebagai hasil dari pemilihan fitur, akurasi klasifikasi yang akan datang dapat ditingkatkan [6].

Manfaat melakukan pemilihan fitur sebelum memodelkan data Anda adalah sebagai berikut:

- a. Mengurangi Overfitting: Data yang lebih sedikit berarti lebih sedikit kesempatan untuk membuat keputusan berdasarkan noise.

- b. Meningkatkan Akurasi: Data yang kurang menyesatkan berarti akurasi pemodelan meningkat.
- c. Mengurangi Kompleksitas: lebih sedikit titik data mengurangi kompleksitas algoritme dan membuatnya lebih mudah dipahami.
- d. Pelatihan Lebih Cepat: Ini memungkinkan algoritme pembelajaran mesin untuk berlatih lebih cepat. 18
- e. Pada sistem ini digunakan dua proses seleksi fitur yaitu proses seleksi fitur sekuensial/forward dan proses seleksi fitur mundur. Pada sistem ini digunakan dua proses seleksi fitur yaitu proses seleksi fitur sekuensial/forward dan proses seleksi fitur mundur

2.2.7 Particle Swarm Optimization (PSO)

Particle swarm optimization adalah salah satu optimasi yang dapat digunakan untuk pengambilan keputusan. PSO adalah teknik optimasi dengan cara menghitung terus menerus calon solusi dengan menggunakan suatu acuan kualitas. PSO mengoptimasi permasalahan dengan cara menggerakkan partikel atau calon solusi di dalam permasalahan menggunakan fungsi tertentu untuk posisi dan kecepatan dari partikel. Pergerakan partikel dipengaruhi oleh solusi terbaik dari partikel tersebut, dan solusi terbaik secara umum yang didapatkan dari partikel lain. Sekumpulan partikel ini dinamakan swarm, swarm ini akan bergerak menuju solusi terbaik.

$$v_{n+1} = v_n + c1rand() * (pbest,n - CurrentPositionn) + c2rand2() * (gbest,n - CurrentPositionn)$$

Particle Swarm Optimization (PSO) adalah teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam Particle Swarm Optimization (PSO), terdapat beberapa teknik untuk optimasi antara lain meningkatkan bobot atribut dari semua atribut

atau variabel yang digunakan, memilih atribut (attribute selection), dan seleksi fitur (Mustopa, 2021). Particle swarm optimization adalah suatu algoritma yang banyak terinspirasi dari perilaku sosial hewan seperti burung, lebah dan ikan. Seekor hewan dalam algoritma PSO akan dianggap sebagai partikel. Partikel ini akan dipengaruhi oleh kecerdasan dari individu hewan itu sendiri dan kecerdasan dari partikel lain dalam satu kelompok. Apabila satu partikel menemukan jalan yang tepat dan terpendek menuju ke suatu sumber makanan, maka yang terjadi adalah partikel-partikel lain tersebut akan mengikuti partikel yang telah menemukan jalan yang tepat dan terpendek tadi [16].

Secara garis besar prosedur PSO dapat dilakukan dalam beberapa langkah.

1. Inisialisasi kecepatan awal bernilai 0 untuk semua partikel seperti pada Persamaan 13.

$$(V_{i,j}(t)=0)$$

$V_{i,j}$ merupakan kecepatan, j adalah letak partikel dan i adalah letak individu dan t adalah iterasi.

2. Inisialisasi posisi awal partikel dengan batasan sesuai range $[x_{min}, x_{max}]$. proses inisialisasi posisi terdapat pada Persamaan 14.

$$x(t)=x_{min}+r(x_{max}-x_{min})$$

X merupakan posisi partikel dan r adalah nilai random

3. Inisialisasi Pbest dan Gbest awal dimana pada iterasi ke 0 nilai Pbest sama dengan posisi awal sesuai dengan Persamaan 15 dan Gbest merupakan Pbest dengan nilai fitness terbaik.

$$(P_{besti,j}(t)=x_{i,j}(t))$$

Pbest merupakan personal best pada individu ke-i dan partikel ke-j. Xij merupakan posisi partikel

4. Update kecepatan dilakukan untuk menentukan arah perpindahan posisi partikel yang ada di populasi. Kecepatan dihitung sesuai Persamaan 16. Terdapat Batasan untuk kecepatan yang digunakan yaitu berdasarkan nilai maksimum dan minimum posisi partikel untuk menentukan batas kecepatan maksimum dan minimum yang dipengaruhi oleh interval (k) yang sebaiknya dilakukan pada proses inisialisasi. Proses update dilakukan seperti pada Persamaan 7 dan 8.

$$t + 1 \quad t \quad t \quad t \quad t \quad t$$

$$v_{i,j} = w \cdot v_{i,j} + c1r1(Pbest_{i,j} - x_{i,j}) + c2r2(Gbest_{g,j} - x_{i,j}) \quad (2.3)$$

$$v_{j,max} = k(x_{j,max} - x_{j,min}) \quad k \in (0,1] \quad (2.4)$$

$$\text{if } v_{ij}(t+1) > v_{maxj} \text{ then } v_{ij}(t+1) = v_{j,max}$$

$$\text{if } v_{ij}(t+1) < -v_{j,max} \text{ then } v_{ij}(t+1) = -v_{j,max}$$

Nilai c1 dan c2 adalah koefisien akselerasi, nilai r1 dan r2 adalah partikel random, nilai w adalah bobot inertia.

5. Update posisi dilakukan untuk menentukan posisi terbaru dari setiap partikel berdasarkan hasil update kecepatan sebelumnya. Setelah didapatkan nilai kecepatan maka dilanjutkan dengan perhitungan sigmoid dari kecepatan tersebut sesuai dengan Persamaan 9 Kemudian hasil sigmoid yang telah didapat akan diproses lebih lanjut pada Persamaan 10 sehingga

didapatkan posisi terbaru Setelah itu menentukan hasil fitness terbaru yang tentunya juga akan mendapat nilai Pbest terbaru.

t

$$sig(vi, j) = 2 + e^{1-vt}, j = 1, 2, \dots, d$$

i, j

$$t \leftarrow t + 1 \text{ if } rand[0,1] > sig(vi, j) \text{ then } x_{i,j} = 0$$

$$t \leftarrow t + 1 \text{ if } rand[0,1] < sig(vi, j) \text{ then } x_{i,j} = 1$$

$$j = 1, 2, \dots, d$$

6. Update Pbest, yaitu dengan membandingkan nilai fitness dari Pbest pada iterasi sebelumnya dengan fitness dari update Posisi. Nilai yang terbaik akan menjadi Pbest yang baru pada iterasi selanjutnya.

$$k = 1 + decimal(s1) \times a^{-1} \quad (21)$$

$$2n1-1$$

2.2.9 Akurasi

Akurasi adalah salah satu metrik untuk mengevaluasi model klasifikasi. Secara informal, akurasi adalah sebagian kecil dari prediksi model kami yang benar. Secara formal, akurasi memiliki definisi sebagai berikut :

$$\text{Akurasi} = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Prediction}}$$

Untuk klasifikasi biner, akurasi juga dapat dihitung dalam hal positif dan negatif sebagai berikut :

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Dimana TP = True

Positif

TN = True

Negatif

FP = False Positif

FN = False Negatif

2.2.9 Confusion Matrix

Matriks konfigurasi adalah tabel yang terdiri dari jumlah baris data uji yang diprediksi benar dan salah dengan model klasifikasi yang digunakan. Tabel Confusion Matrix diperlukan untuk memilih kinerja terbaik dari sebuah model klasifikasi [17].

2.2.10 Kurva ROC dan AUC

Dalam Machine Learning, pengukuran kinerja adalah tugas penting. Jadi dalam masalah klasifikasi, kita dapat mengandalkan Kurva AUC - ROC. Ketika kita perlu memeriksa atau memvisualisasikan kinerja masalah klasifikasi multi-kelas, kita menggunakan kurva AUC (Area Under The Curve) ROC (Receiver Operating Characteristics). Ini adalah salah satu metrik evaluasi terpenting untuk memeriksa kinerja model klasifikasi apa pun. Itu juga ditulis sebagai AUROC (Area Di Bawah Karakteristik Operasi Penerima) [18].

Area di Bawah Kurva (AUC) adalah ukuran dari kesesuaian metode. AUC mewakili nilai sensitivitas dan spesifisitas dengan nilai batas 0 sampai 1 . Selanjutnya, hasil evaluasi ini dikategorikan berdasarkan nilai yang diperoleh dari setiap pengukuran [19]. Gorunescu [20] mengkategorikan hasil klasifikasi berdasarkan pada nilai AUC sebagai berikut:

- $0,90 - 1,00$ = klasifikasi sangat baik;
- $0,80 - 0,90$ = klasifikasi baik;
- $0,70 - 0,80$ = klasifikasi wajar;
- $0,60 - 0,70$ = klasifikasi buruk;
- $0,50 - 0,60$ = gagal.