



Prediksi Loyalitas Karyawan Dengan Menggunakan Fitur Optimasi Pembobotan PSO dan Algoritma Decision Tree C4.5

Muhamad Fahrurrozi*¹, Sriyanto²

^{1,2,3} Fakultas Ilmu Komputer, Informatics & Business Institute Darmajaya, Jl. 2.A. Pagar AlamNo. 93, Bandar Lampung - Indonesia 35142; Telp. (0721) 787214 Fax. (0721) 700261

*Email Penulis Korespondensi: *¹fahrurrozi.muhamad2@gmail.com

Abstrak

Salah satu elemen dalam perusahaan yang sangat penting adalah Sumber Daya Manusia (SDM). Pengelolaan SDM dari suatu perusahaan sangat mempengaruhi banyak aspek penentu keberhasilan kerja dari perusahaan tersebut. Aset terbesar yang dimiliki sebuah perusahaan bukanlah miliknya bisnis atau tempat atau hal lain, tapi itu karyawan. Pada dasarnya perusahaan menginginkan karyawannya dapat bekerja secara loyalitas. Banyak sekali factor-faktor yang mempengaruhi kinerja seseorang karyawan, seperti umur, jenis kelamin, tingkat kedisiplinan, dll. Kebutuhan dari perusahaan untuk mengetahui faktor utama yang mempengaruhi loyalitas karyawan dengan menggunakan data mining. Untuk bisa mendapatkan informasi dari data-data yang ada perlu dilakukan proses data mining seperti klasifikasi. Pada permasalahan diatas mengenai karyawan maka untuk dapat memprediksi loyalitas karyawan dalam suatu perusahaan dengan mempertimbangkan berbagai faktor dan menggunakan algoritma untuk mendapatkan prediksi dengan akurasi yang tinggi salah satunya adalah Decision Tree C4.5 untuk mengklasifikasikan faktor paling penting untuk loyalitas karyawan. Pengujian sebelumnya oleh Sai Chandan P Redgg pada tahun 2021 menghasilkan akurasi yang cukup tinggi yaitu sebesar 86.39%. Merujuk hasil akurasi dari penelitian tersebut dapat dilihat bahwa Decision Tree menghasilkan akurasi yang tinggi namun hasil akurasi tersebut masih dapat ditingkatkan lagi dengan melakukan penelitian lanjutan untuk menghasilkan akurasi lebih tinggi dengan menambahkan optimization feature weighting (PSO).

Kata kunci— Loyalitas Karyawan, Klasifikasi, Algoritma Decision Tree C4.5, PSO.

Abstract

One of the most important elements in a company is Human Resources (HR). HR management of a company greatly influences many aspects of determining the success of the work of the company. The greatest asset a company has is not its business or place or anything else, it's its employees. Basically the company wants its employees to work loyally. There are so many factors that affect the performance of an employee, such as age, gender, level of discipline, etc. The need for companies to know the main factors that influence employee loyalty by using data mining. To be able to get information from existing data, it is necessary to carry

out data mining processes such as classification. In the above problems regarding employees, to be able to predict employee loyalty in a company by considering various factors and using algorithms to obtain predictions with high accuracy, one of them is Decision Tree C4.5 to classify the most important factors for employee loyalty. The previous test by Sai Chandan P Redgg in 2021 resulted in a fairly high accuracy of 86.39%. Referring to the results of the accuracy of the research, it can be seen that the Decision Tree produces high accuracy, but the results of this accuracy can still be improved by conducting further research to produce higher accuracy by adding optimization feature weighting (PSO).

Keywords— Employee Loyalty, Classification, Decision Tree Algorithm C4.5, PSO.

1. PENDAHULUAN

Salah satu elemen dalam perusahaan yang sangat penting adalah Sumber Daya Manusia (SDM). Pengelolaan SDM dari suatu perusahaan sangat mempengaruhi banyak aspek penentu keberhasilan kerja dari perusahaan tersebut. Jika SDM dapat diorganisir dengan baik, maka diharapkan perusahaan dapat menjalankan semua proses usahanya dengan baik [1]. Karyawan yang baik adalah selalu menjadi aset berharga bagi perusahaan, dan kepergian mereka dapat mengakibatkan berbagai masalah, termasuk kerugian finansial, kinerja keseluruhan yang buruk, dan hilangnya akumulasi keahlian. Selanjutnya, dibandingkan dengan merekrut karyawan baru, perusahaan secara signifikan mengeluarkan biaya yang mahal dan memakan waktu [2].

Pada dasarnya perusahaan menginginkan karyawan nya dapat bekerja secara maksimal sesuai dengan keahliannya masing-masing. Banyak sekali faktor-faktor yang mempengaruhi kinerja seseorang karyawan, seperti umur, jenis kelamin, tingkat kedisiplinan, dll [3]. Menurut [4], kinerja karyawan adalah dipengaruhi oleh psikologis, lingkungan, sosial, dan faktor fisik. Kebutuhan dari perusahaan untuk mengetahui faktor utama yang mempengaruhi loyalitas karyawan dengan menggunakan data *mining*.

Untuk bisa mendapatkan informasi dari data-data yang ada perlu dilakukan proses data *mining* seperti klasifikasi. Klasifikasi merupakan sebuah proses untuk menentukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui, di dalam klasifikasi juga diberikan sejumlah *record* yang dinamakan training set, yang terdiri dari beberapa atribut, atribut dapat berupa kontinu ataupun kategoris, salah satu atribut menunjukkan kelas untuk *record* [5].

Pada permasalahan diatas mengenai karyawan maka untuk dapat mengatasi memprediksi loyalitas suatu karyawan dalam suatu perusahaan dengan mempertimbangkan berbagai faktor dan menggunakan algoritma untuk mendapatkan prediksi dengan akurasi yang besar salah satunya adalah *Decision Tree* C4.5 untuk mengklasifikasikan faktor paling penting untuk loyalitas karyawan. Pengujian sebelumnya oleh Sai Chandan P Redgg pada tahun 2021 menghasilkan akurasi yang cukup tinggi yaitu sebesar 86.39%. Merujuk hasil akurasi dari penelitian tersebut dapat dilihat bahwa *Decision Tree* menghasilkan akurasi yang tinggi namun hasil akurasi tersebut masih dapat ditingkatkan lagi dengan melakukan penelitian lanjutan untuk menghasilkan akurasi lebih tinggi dengan menambahkan *optimization feature weighting* (PSO). dan menggunakan dataset *kaggle*. Berdasarkan hasil penelitian yang dilakukan peneliti sebelumnya maka masih perlu dikembangkan lagi agar klasifikasi prediksi loyalitas karyawan mendapatkan tingkat akurasi yang lebih tinggi .

2. METODE PENELITIAN

2.1 Jenis Penelitian

Dalam penelitian ini terdapat dua pendekatan utama, yaitu pendekatan kualitatif dan pendekatan kuantitatif. Pendekatan kualitatif digunakan untuk menganalisa kajian literatur yang berkenaan dengan variabel-variabel yang digunakan dalam pengumpulan data [6].

Sedangkan pendekatan kuantitatif merupakan metode penelitian yang digunakan sebagai penelitian populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian dan analisis data bersifat statistik atau kuantitatif dengan tujuan untuk menguji hipotesis yang telah ditetapkan [7].

2.2 Metode Pengumpulan Data

Pada tahap ini metode pengumpulan data dibagi menjadi dua sumber data yakni data primer dan data skunder. Adapun data primer yaitu data yang dikumpulkan dari sumbernya langsung. Sedangkan data skunder yaitu data yang dikumpulkan dari peneliti sebelumnya artinya peneliti tidak harus mengambil datanya langsung ke lapangan. Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang akan digunakan adalah dengan metode pengumpulan data skunder. Data dari penelitian ini diambil dari situs *kaggle dataset repository* (<https://www.kaggle.com/datasets/tejashvi14/employee-future-prediction>) sedangkan untuk data pendukung didapatkan dari buku jurnal dan publikasi lainnya.

2.3 Validasi Data

Validasi data dimana data penelitian ini diambil dari *Kaggle Dataset* yang terdiri 9 atribut dimana terdiri dari 8 atribut Fitur dan 1 atribut label. Pada dataset tersebut dilakukan *validasi* data atau *preprocessing* dengan cara menghilangkan data *missing value*. Data yang didapat dari situs *Kaggle Dataset* sebanyak 4.653 *record*, dari data tersebut dilakukan *preprocessing* tetapi tidak ada data yang *redundance* atau *missing value* sehingga data yang digunakan tetap sejumlah 4.653 *record*.

TABEL I
ATRIBUT, TYPE DATA DAN NILAI KATEGORI DALAM
KLASIFIKASI LOYALITAS KARYAWAN

Atribut	Deskripsi	Kriteria
Education	Sekolah	Bachelors
		Masters
		PHD
JoiningYear	Tahun Bergabung	2012-2018
City	Kota	Bangalore
		Pune
		New Delhi
PaymentTier	Range Gaji	1= Highest (Tinggi)
		2= Mid Level (Sedang)
		3= Lowest (Rendah)
Age	Umur	22-41
Gender	Jenis Kelamin	Male = Laki-laki
		Female = Perempuan
EverBenched	Tidak pernah keluar dai proyek selama 1 bulan	Yes = Iya
		No = Tidak
ExperienceInCurrentDomain	Pengalaman di lapangan saat ini	0-7 (paling tinggi 7)

LeaveOrNot	Apakah karyawan tinggalkan perusahaan dalam 2 tahun mendatang	0 = No
		1 = Yes

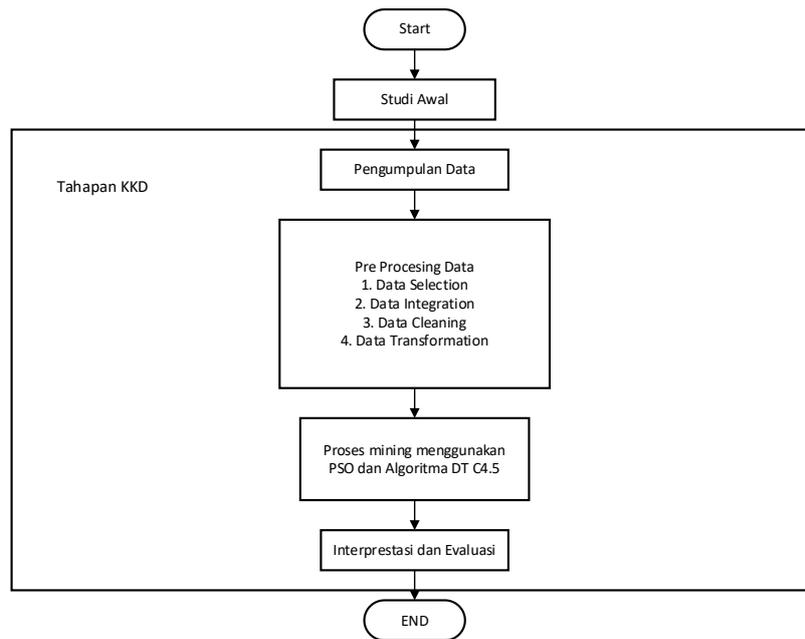
Berikut merupakan deskripsi dari atribut diatas :

1. *Education* (Pendidikan)
Education merupakan atribut yang mencakup Pendidikan/sekolah terakhir untuk loyalitas karyawan dengan kategori *Bachelors, masters, PHD*.
2. *Joining Year* (Tahun Masuk)
Joining Year merupakan atribut yang mencakup tahun masuk/bergabung ke perusahaan dengan kategori tahun dari 2012-2018.
3. *City* (Kota)
City merupakan atribut yang mencakup kota dari pegawai dataset ini memiliki 3 data kota yaitu (Bangalore, Pune, New Delhi).
4. *Payment Tier* (Range Gaji)
Payment Tier merupakan atribut yang mencakup *range* gaji pegawai untuk menentukan loyalitas pegawai dalam perusahaan yang memiliki kategori:
 - a) 1 = *Highest* (tinggi)
 - b) 2 = *Mid Level* (sedang)
 - c) 3 = *Lowlest* (rendah)
5. *Age* (Umur)
Age merupakan atribut yang mengkategorikan umur pegawai untuk sebuah loyalitas di perusaan, data yang diambil dari *dataset* yaitu umur 22-41 tahun.
6. *Gender* (Jenis Kelamin)
Gender merupakan atribut yang mengkategorikan jenis kelamin sebuah pegawai dalam loyalitas karyawan. Jenis kelamin ini dibagi menjadi 2 yaitu, *male* = laki-laki dan *female* = perempuan.
7. *Ever Benched* (Tidak pernah keluar dai proyek selama 1 bulan)
Ever Benched merupakan atribut yang mengkategorikan sebuah pegawai atau karyawan tidak pernah keluar dari sebuah proyek selama kurun waktu 1 bulan. Kategori dari atribut ini diantaranya *Yes* dan *No*.
8. *ExperienceInCurrentDomain* (Pengalaman di lapangan saat ini)
ExperienceInCurrentDomain merupakan atribut yang mencakup pengalaman pegawai/raryawan. Pada kategori ini semakin tinggi nilainya maka pengalaman pegawai tersebut semakin baik. Dimana kategori tersebut meliputi nilai 0-7.
9. *LeaveOrNot* (Apakah karyawan tinggalkan perusahaan)
LeaveOrNot merupakan atribut label yang yang menandakan pegawai/karyawan tersebut dalam kurun waktu 2 tahun keluar atau tidak dengan kategori 0=*No* dan 1=*Yes*.

2. 4 Kerangka Pemikiran

Penelitian ini terdiri dari beberapa tahapan. Permasalahan pada penelitian ini yaitu membandingkan penelitian sebelumnya mengenai klasifikasi loyalitas karyawan dengan menggunakan algoritma yang sama yaitu *Decision Tree C4.5* dan ditambahkan metode PSO untuk mendapatkan akurasi yang tinggi.

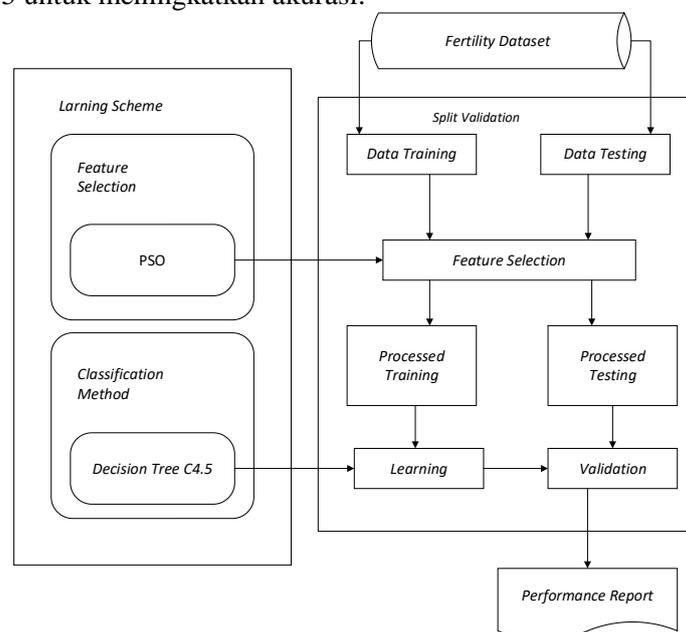
Atas dasar alasan tersebut, maka dilakukan penelitian menggunakan metode klasifikasi algoritma *Decision Tree C4.5* dan PSO dalam memprediksi loyalitas karyawan. Pengujian metode dilakukan dengan cara *confusion matrix* dan kurva ROC, menggunakan tools Rapidminer Studio v.9. berikut dibawah ini kerangka pemikiran yang penulis buat untuk penelitian ini:



Gambar 1. Kerangka Pemikiran

2. 5 Proses PSO

Proses mining menggunakan metode *Particle Swam Optimization* (PSO) dan ALgoritma *Decision Tree C4.5* untuk meningkatkan akurasi.



Gambar 2. Alur Proses PSO

- Tahapan pertama yaitu pengumpulan *dataset fertility* yang diambil dari *Kaggle repository*.
- Berikutnya pembagian data menggunakan *split validation*, yaitu membagi data menjadi *data training* dan *data testing*.
- Pada tahap *learning scheme* menggunakan *Particle Swarm Optimization* untuk optimasi, sehingga akan mengoptimasi nilai bobot pada setiap atribut.

Pengujian PSO

- d. Selanjutnya proses klasifikasi menggunakan algoritma *Decision Tree* C4.5 dan masuk ketahap *validasi* atau melihat kebenaran dan hasil dari proses sebelumnya.

2. 6 Data Mining

Data mining merupakan suatu istilah yang dapat digunakan untuk menguraikan penemuan pengetahuan di dalam suatu *database*. *Data mining* juga adalah sebuah 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 [8].

2. 7 Algoritma Decision Tree C4.5

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 [9].

C4.5 merupakan salah satu solusi pemecahan kasus yang sering digunakan dalam pemecahan masalah pada teknik klasifikasi [10]. Pemilihan atribut sebagai simpul, baik simpul akar (*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Penghitungan nilai *Gain* digunakan rumus Persamaan 1.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

S : Himpunan kasus

A : Atribut

n : Jumlah partisi himpunan atribut A

$||$: Jumlah kasus pada partisi ke- i

$|S|$: Jumlah kasus dalam S

Untuk menghitung nilai *Entropy* dapat dilihat pada Persamaan 2

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2(p_i) \quad (2)$$

n : Jumlah partisi S

p_i : Proporsi dari terhadap S

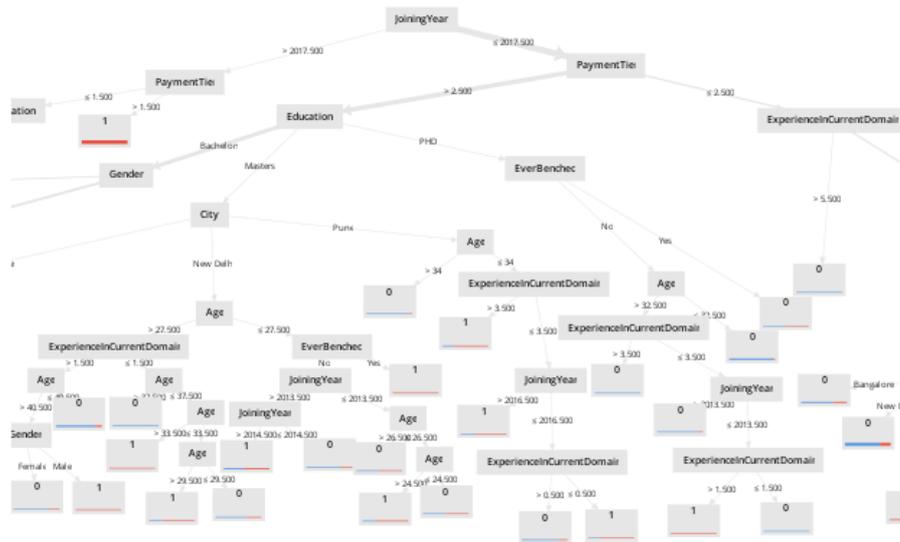
3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* C4.5 dan PSO untuk mendapatkan tingkat akurasi klasifikasi loyalitas karyawan. Hasil dari penelitian ini berupa pengolahan data kualitatif dan data kuantitatif dengan perhitungan yang dilakukan pada sebuah dataset publik.

3.1 Pemodelan Algoritma Decision Tree C4.5

Pemodelan algoritma *Decision Tree* C4.5 terdiri dari 11 atribut yang merupakan atribut dari prediksi loyalitas karyawan dan class yang merupakan hasil akhir prediksi.

Model dari algoritma *Decision Tree* C4.5 yaitu berupa pohon keputusan, agar lebih mudah dalam membuat pohon keputusan, langkah pertama adalah menghitung jumlah class yang keluar dan tidak keluar dari perusahaan, masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*.



Gambar 3. Model Pohon Keputusan Klasifikasi Loyalitas Karyawan Menggunakan Algoritma C4.5

Pada pemodelan pohon keputusan diatas puncak paling atas adalah *joining year* (tahun masuk) pada perusahaan. Ini adalah top atribut paling penting dalam klasifikasi loyalitas karyawan pada *dataset* tersebut.

3.2 Hasil Akurasi Algoritma C4.5

Pada eksperimen algoritma yang telah dilakukan maka didapatkan hasil untuk prediksi loyalitas karyawan sebesar 87.37%. hasil akurasi dapat dilihat dari gambar berikut:

accuracy: 87.37%

	true 0	true 1	class precision
pred. 0	2357	385	85.96%
pred. 1	85	895	91.33%
class recall	96.52%	69.92%	

Gambar 4. Hasil Akurasi Klasifikasi Loyalitas Karyawan Menggunakan Algoritma C4.5

3.3 Hasil Akurasi Algoritma C4.5 dan PSO

Pada eksperimen algoritma C4.5 dan PSO yang telah dilakukan maka didapatkan hasil untuk prediksi loyalitas karyawan sebesar 88.80%. hasil akurasi dapat dilihat dari gambar berikut:

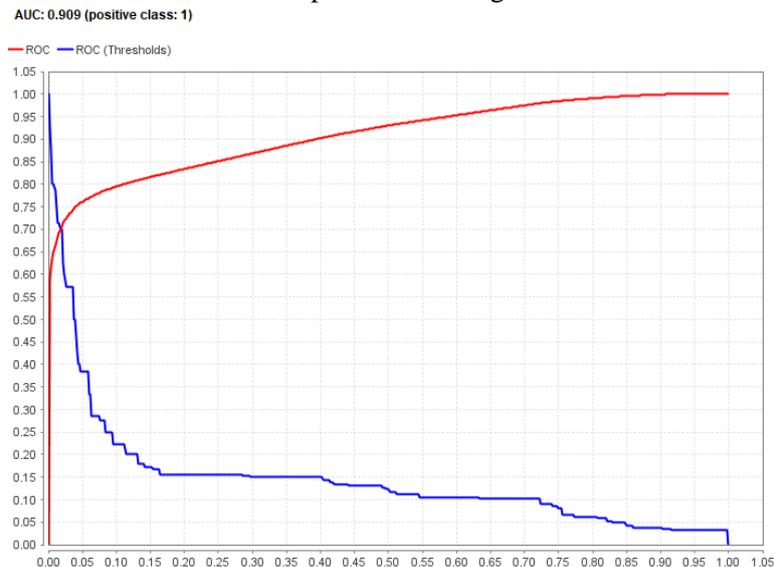
accuracy: 88.80%

	true 0	true 1	class precision
pred. 0	2379	354	87.05%
pred. 1	63	926	93.63%
class recall	97.42%	72.34%	

Gambar 5. Hasil Akurasi Klasifikasi Loyalitas Karyawan Menggunakan Algoritma C4.5 dan PSO

3.4 Hasil Visualisasi ROC Curve (AUC)

Hasil dari prediksi loyalitas karyawan dapat dilihat dalam kurva ROC (*Receiver Operating Characteristic*). Hasil dari kurva ROC dapat dilihat dari gambar berikut:

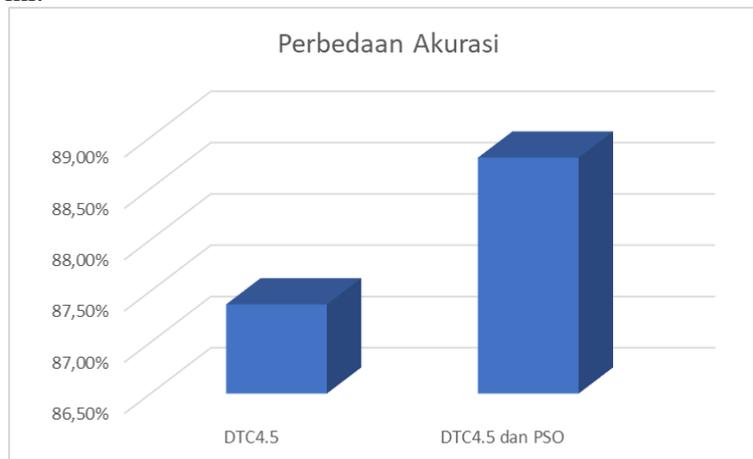


Gambar 6. Hasil Visualisasi ROC Curve (AUC)

Pada gambar dibawah ini dijelaskan bahwa nilai AUC:0,909 (*positive class :1*) garis merah merupakan ROC dan garis biru merupakan ROC (*Thresholds*).

3.5 Perbandingan akurasi C4.5 dan C4.5 + PSO

Berikut ini adalah hasil perbandingan sebelum dan sesudah antara algoritma C4.5 dan C4.5 + PSO, dari perbandingan berikut ini dapat dilihat bahwa ada peningkatan akurasi yang sebelumnya hanya 87.37% ke 88.80%. perbedaan peningkatan akurasi ini dapat dilihat dari gambar berikut ini.



Gambar 7. Perbandingan Akurasi C4.5 dan C4.5 + PSO

4. KESIMPULAN

- Dari penjelasan diatas ini dapat diambil kesimpulan penelitian dengan menggunakan Decision tree C4.5 dan PSO dilakukan agar dapat mengetahui hasil atribut yang paling berpengaruh dan meningkatkan nilai akurasi yang dihasilkan.

- Dari 9 atribut yang terdapat diatas, kemudian atribut tersebut selanjutnya diseleksi mana atribut terpenting yang mempengaruhi loyalitas karyawan sehingga dapat disimpulkan atribut paling penting adalah *joining year*.
- Pada penelitian sebelumnya hanya menggunakan algoritma *Decision Tree* sehingga akurasi yang didapatkan kurang maksimal, sehingga dilakukannya peningkatan akurasi dengan menambahkan metode PSO. dari hasil yang dapat dilihat DTC4.5 menghasilkan akurasi 87.37%. dan setelah ditambahkan PSO mengalami peningkatan akurasi 88.80%. maka dari itu metode PSO ini cukup baik di gabungkan dengan algoritma *Decision Tree C4.5* karna meningkatkan nilai akurasi yang cukup baik.

5. SARAN

- Dataset *public* tidak terdapat error sehingga dataset asli yang digunakan.
- Lakukan perbandingan analisis akurasi dengan algoritma yang lain dengan menggunakan dataset yang sama untuk mengetahui perbandingan nilai akurasinya.
- Untuk mengimprove akurasi dapat mengubah parameter sampling linier disesuaikan dengan dataset.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] W. Wahyono and A. Nugroho, "Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kompetensi Karyawan Pt Multistrada Arah Sarana," *Joutica*, vol. 3, no. 1, p. 145, 2018, doi: 10.30736/jti.v3i1.201.
- [2] H. Mallafi and D. H. Widyantoro, "Prediction modelling in career management," *Proc. - Cybern. 2016 Int. Conf. Comput. Intell. Cybern.*, vol. 6, no. 11, pp. 17–21, 2017, doi: 10.1109/CyberneticsCom.2016.7892560.
- [3] T. Novianti and I. Santosa, "PENENTUAN JADWAL KERJA BERDASARKAN KLASIFIKASI DATA KARYAWAN MENGGUNAKAN METODE DECISION TREE C4.5 (Studi Kasus Universitas Muhammadiyah Surabaya)," *J. Komunika J. Komunikasi, Media dan Inform.*, vol. 5, no. 1, p. 1, 2016, doi: 10.31504/komunika.v5i1.633.
- [4] et al. Fajarwati, "Pengaturan sistem shift kerja untuk meningkatkan performance serta mengurangi keluhan karyawan," *J. Teknol. Technoscientia*, vol. 4, no. 1, pp. 37–45, 2011.
- [5] A. Som, "Computational Biology and Chemistry," *Computer (Long. Beach. Calif.)*, vol. 33, pp. 373–378, 2009.
- [6] A. Fauzi, B. Suharjo, and M. Syamsun, "Pengaruh Sumber Daya Finansial, Aset Tidak Berwujud dan Keunggulan Bersaing yang Berimplikasi Terhadap Kinerja Usaha Mikro, Kecil dan Menengah di Lombok NTB," *Manaj. IKM J. Manaj. Pengemb. Ind. Kecil Menengah*, vol. 11, no. 2, pp. 151–158, 2017, doi: 10.29244/mikm.11.2.151-158.
- [7] A. Syaifullah, "Analisis pengaruh financial leverage dan operating leverage terhadap stock return," *Inovasi*, vol. 14, no. 2, p. 53, 2018, doi: 10.29264/jinv.v14i2.1928.
- [8] E. Turban and dkk., *Decicion Support Systems and Intelegent Systems*. Andi Offset, 2005.
- [9] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innov. Res. Informatics*, vol. 1, no. 2, pp. 63–69, 2019, doi: 10.37058/innovatics.v1i2.918.
- [10] D. Nofriansyah and G. W. Nurcahyo, *Algoritma Data Mining Dan Pengujian*. Yogyakarta: Deepublish, 2019.