



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 1 Tahun 2025 Page 1712-1726

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Perbandingan Algoritma Naïve Bayes Dan C4.5 Dalam Prediksi Keputusan Karyawan Untuk Meninggalkan Perusahaan

Ferdinan Daewana^{1✉}

Universitas Mercu Buana

Email: ferdinandaewana22@gmail.com^{1✉}

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan C4.5 dalam memprediksi keputusan karyawan terkait kemungkinan mereka meninggalkan perusahaan. Data yang digunakan terdiri dari data primer yang diperoleh melalui wawancara dengan HRD di CV. Dunia Pustaka dan data sekunder yang diunduh dari Kaggle. Penelitian ini mengimplementasikan kedua algoritma menggunakan bahasa pemrograman Python untuk melakukan analisis klasifikasi dan mengukur kinerja model berdasarkan metrik akurasi, precision, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa algoritma C4.5 memiliki kinerja yang lebih baik dibandingkan Naïve Bayes. C4.5 mencapai akurasi 92%, precision 100%, recall 90.24%, dan F1-score 95.84%, sementara Naïve Bayes hanya memperoleh akurasi 84%, precision 94.59%, recall 85.37%, dan F1-score 89.6%. Perbandingan ini menunjukkan bahwa C4.5 lebih efektif dalam mengklasifikasikan data dan memberikan prediksi yang lebih akurat terkait keputusan karyawan untuk meninggalkan perusahaan. Penelitian ini memberikan wawasan penting bagi perusahaan dalam memahami faktor-faktor yang mempengaruhi keputusan karyawan serta cara-cara untuk mempertahankan karyawan. Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan metode analisis lainnya untuk meningkatkan akurasi prediksi dan memahami lebih dalam faktor-faktor yang mempengaruhi turnover karyawan.

Kata Kunci: *Naïve Bayes, C4.5, Klasifikasi, Prediksi Keputusan Karyawan, Turnover, Akurasi, Precision, Recall, F1-Score*

Abstract

This research aims to compare the performance of the Naïve Bayes and C4.5 algorithms in predicting employee decisions regarding the likelihood of leaving the company. The data used includes primary data obtained through interviews with the HRD at CV. Dunia Pustaka and secondary data downloaded from Kaggle. Both algorithms were implemented using Python programming language for classification analysis, with model performance evaluated based on accuracy, precision, recall, and F1-score metrics. The evaluation results indicate that the C4.5 algorithm outperforms Naïve Bayes. C4.5 achieved an accuracy of 92%, precision of 100%, recall of 90.24%, and F1-score of 95.84%, while Naïve Bayes reached an accuracy of 84%, precision of 94.59%, recall of 85.37%, and F1-score of 89.6%. This comparison shows that C4.5 is more effective in classifying data and providing more accurate predictions of employee decisions to leave the company. The research offers valuable insights for companies to understand the factors influencing employee decisions and strategies for retaining employees. Future studies are recommended to use larger datasets and other analytical methods to improve prediction accuracy and gain a deeper understanding of the factors affecting employee turnover.

Keywords: *Naïve Bayes, C4.5, Classification, Employee Decision Prediction, Turnover, Accuracy, Precision, Recall, F1-Score*

PENDAHULUAN

Tingkat pergantian karyawan yang tinggi merupakan tantangan utama bagi banyak perusahaan, karena kehilangan karyawan dengan kontribusi besar dapat berdampak negatif pada operasional dan pencapaian tujuan perusahaan (Putri, 2024). Pergantian karyawan tidak hanya menciptakan kekosongan dalam tenaga kerja, tetapi juga membebani perusahaan secara finansial. Biaya yang diperlukan untuk menggantikan seorang karyawan dapat melampaui 100% dari gaji tahunannya. Selain itu, turnover dapat mengganggu produktivitas tim dan mengurangi potensi pendapatan perusahaan (Stephen, 2022).

Proses rekrutmen, seleksi, dan pelatihan yang dilakukan oleh tim Sumber Daya Manusia (HR) membutuhkan investasi waktu, tenaga, dan biaya yang signifikan untuk memastikan karyawan mampu menjalankan tugas dengan optimal (Utami, no date). Oleh karena itu, menjaga keberadaan tenaga kerja yang berkualitas menjadi hal yang krusial bagi organisasi (RAMADHANI, 2023). Tim yang solid dan kompeten dapat menciptakan lingkungan kerja yang mendukung serta berkontribusi terhadap keberlanjutan dan pertumbuhan perusahaan dalam jangka panjang. Untuk mengatasi permasalahan tersebut, peneliti berencana membandingkan dua metode klasifikasi, yaitu Naïve Bayes dan C4.5. Tujuan dari perbandingan ini adalah untuk mengevaluasi tingkat akurasi kedua algoritma. Algoritma C4.5 digunakan untuk membentuk pohon keputusan, sedangkan Naïve Bayes

merupakan metode klasifikasi yang didasarkan pada penerapan teorema Bayes melalui perhitungan sejumlah probabilitas(Zees, 2023).

Beberapa penelitian sebelumnya juga telah membahas topik serupa, salah satunya adalah penelitian berjudul "*Predicting Employee Attrition Using Machine Learning Techniques*". Penelitian ini menguji berbagai algoritma klasifikasi, termasuk Naive Bayes untuk model multivariat Bernoulli, Logistic Regression, K-Nearest Neighbors (K-NN), Decision Tree, Random Forest, Support Vector Machines (SVM), Linear Support Vector Machines (LSVM), serta Gaussian Naïve Bayes (Tulus, Hadiana and Santikarama, 2022). Hasilnya menunjukkan bahwa algoritma Gaussian Naïve Bayes mencapai recall rate sebesar 0,54, false negative rate 4,5%, dan akurasi sebesar 82,5%. Di sisi lain, akurasi tertinggi diperoleh dari klasifikasi LSVM dengan angka 87,9%. Penelitian tersebut mengidentifikasi faktor utama yang memengaruhi attrisi karyawan, yakni pendapatan bulanan, usia, lembur, dan jarak tempat tinggal dari kantor (Maahiroh, 2024)

Penelitian berjudul "Perbandingan Metode Seleksi Fitur untuk Mengoptimasi Model Support Vector Machine dalam Memprediksi Turnover Pegawai" mengevaluasi efektivitas metode seleksi fitur pada model Support Vector Machine (SVM). Hasilnya menunjukkan bahwa seleksi fitur dapat memengaruhi hasil evaluasi model, baik secara positif maupun negatif, bergantung pada metode yang digunakan (Andini and Hardiyanti, 2024). Metode wrapper terbukti meningkatkan performa model hingga mencapai 60%, sedangkan metode filter dan embedded justru menurunkan performa masing-masing hingga 55%. Variasi hasil ini kemungkinan besar disebabkan oleh perbedaan fitur yang dipilih oleh masing-masing metode seleksi(AMIR, 2018).

Penelitian lain dimana kedua penelitian tersebut sama-sama membahas prediksi attrisi karyawan dengan memanfaatkan teknik machine learning, namun terdapat perbedaan dalam pendekatan yang diterapkan serta hasil yang diperoleh(Francesca Fallucchi, 2020). Penelitian ini mengevaluasi berbagai algoritma klasifikasi dan menemukan bahwa Gaussian Naïve Bayes menghasilkan kinerja terbaik dengan recall rate sebesar 0,54 dan akurasi mencapai 82,5%(Fadel *et al.*, 2024). Di sisi lain, LSVM mencatat akurasi tertinggi sebesar 87,9%(Handayani, 2017). Faktor-faktor utama yang memengaruhi tingkat attrisi karyawan dalam studi ini meliputi pendapatan bulanan, usia, lembur, serta jarak tempat tinggal dari kantor. Penelitian yang dilakukan difokuskan pada pengoptimalan model SVM melalui penerapan metode seleksi fitur. Hasil penelitian menunjukkan bahwa metode *wrapper* berhasil meningkatkan performa model hingga mencapai 60%, sementara metode *filter* dan *embedded* justru menurunkan performa menjadi 55% (HUMAIROSI, 2022).

Perbedaan mendasar antara penelitian ini dengan studi yang dilakukan oleh Fallucchi et al. terletak pada fokus penelitian; Fallucchi et al. lebih menekankan evaluasi berbagai algoritma, sedangkan penelitian Abiyyu menyoroti pengaruh seleksi fitur terhadap performa model SVM. Berdasarkan latar belakang permasalahan dan beberapa penelitian sebelumnya yang telah dibahas, penulis memutuskan untuk menerapkan perbandingan antara algoritma Naïve Bayes dan C4.5 dalam prediksi keputusan karyawan. Perbandingan ini bertujuan untuk mengetahui tingkat keakuratan dari kedua algoritma tersebut, sehingga dapat membantu manajemen sumber daya manusia dalam mengambil keputusan yang lebih tepat dan efektif (Rianti and Andarsyah, 2024).

METODE PENELITIAN

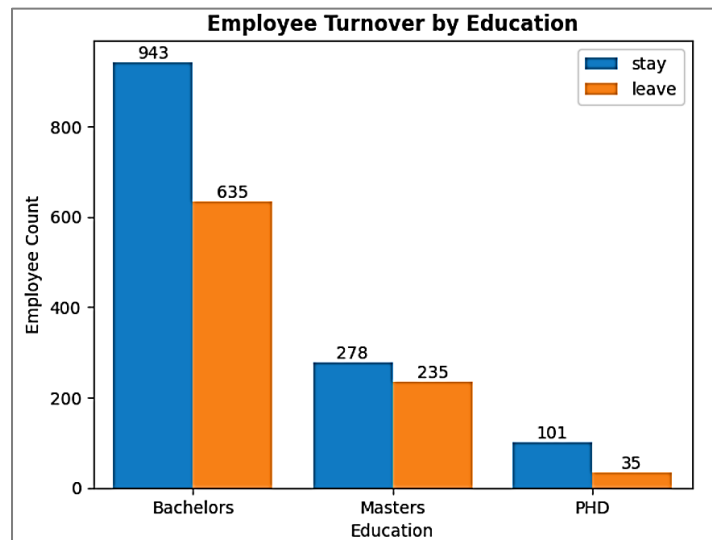
Penelitian ini menggunakan pendekatan kuantitatif, yang fokus pada pengumpulan dan analisis data numerik untuk menjawab pertanyaan penelitian secara terukur. Pendekatan ini dipilih untuk memperoleh data objektif yang dapat dianalisis secara statistik. Tujuan penelitian adalah untuk membandingkan kinerja dua algoritma, Naïve Bayes dan C4.5, dalam memprediksi keputusan karyawan untuk meninggalkan perusahaan. Keunggulan utama pendekatan ini adalah kemampuannya menghasilkan data yang dapat diukur secara numerik dan mendukung generalisasi hasil ke populasi lebih luas. Langkah penelitian dimulai dengan identifikasi variabel, yaitu algoritma sebagai variabel independen dan keputusan karyawan sebagai variabel dependen. Data yang dikumpulkan bersifat numerik, mencakup faktor-faktor seperti kepuasan kerja dan gaji. Kedua algoritma diterapkan pada dataset yang telah diproses, dan kinerja diuji menggunakan metrik seperti akurasi dan F1-score. Selanjutnya, uji statistik dilakukan untuk mengevaluasi perbedaan signifikan antara kinerja kedua algoritma.

HASIL DAN PEMBAHASAN

Berikut merupakan bentuk dari visualisasi menggunakan pemrograman *Phyton* pada *Google Colabulatory* beberapa atribut yaitu sebagai berikut :

1. Visualisasi Atribut *Education* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisai atribut *Education* berdasarkan keputusan karyawan menggunakan *Tools Google Colabulatory*.

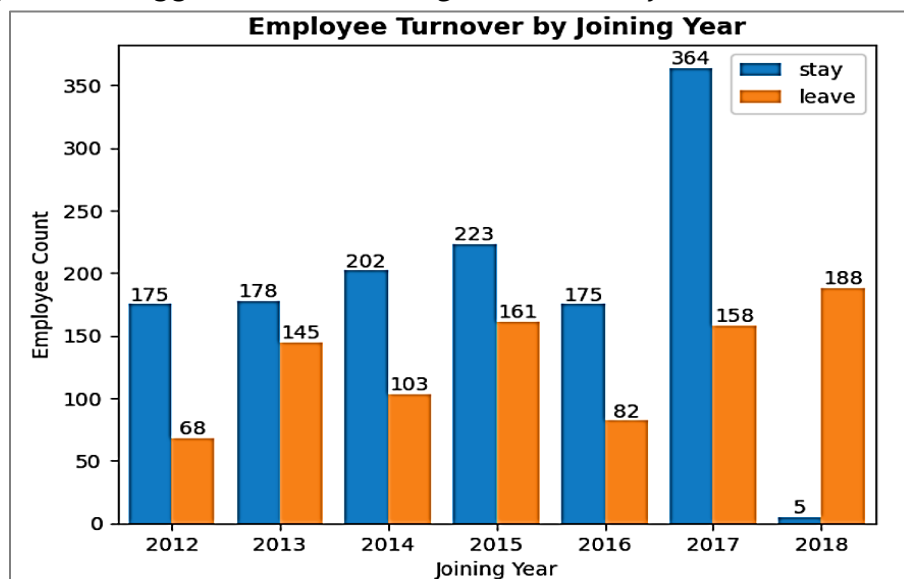


Gambar 4. 1 Visualisasi Atribut Education

Gambar 4.6 adalah visualisasi dari atribut *Education* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 943 karyawan yang tinggal dan 635 karyawan yang meninggalkan perusahaan dengan pendidikan "*Bachelors*", 278 karyawan yang tinggal dan 235 karyawan yang meninggalkan perusahaan dengan pendidikan "*Master*" dan 101 karyawan yang tinggal dan 35 karyawan yang meninggalkan perusahaan dengan pendidikan "*PHD*".

2. Visualisasi Atribut *Joining Year* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisai atribut *Joining Year* berdasarkan keputusan karyawan menggunakan *Tools Google Colabatory*.



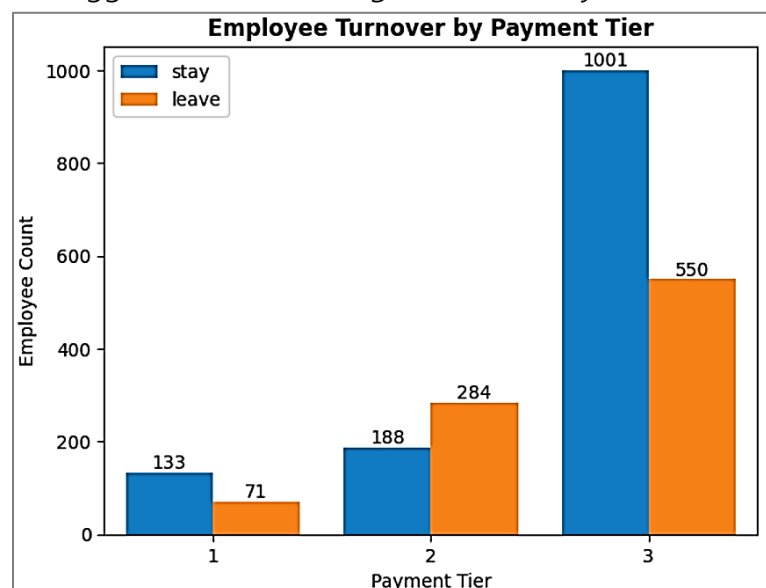
Gambar 4. 2 Visualisasi Atribut Joining Year

Gambar 4.7 adalah visualisasi dari atribut *Joining Year* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 175 karyawan yang tinggal

dan 68 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2012", 178 karyawan yang tinggal dan 145 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2013", 202 karyawan yang tinggal dan 103 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2014", 223 karyawan yang tinggal dan 161 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2015", 175 karyawan yang tinggal dan 82 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2016", 364 karyawan yang tinggal dan 158 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2017" dan 5 karyawan yang tinggal dan 188 karyawan yang meninggalkan perusahaan dengan tahun bergabung "2018".

3. Visualisasi Atribut *Payment Tier* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisasi atribut *Payment Tier* berdasarkan keputusan karyawan menggunakan *Tools Google Colabulatory*.

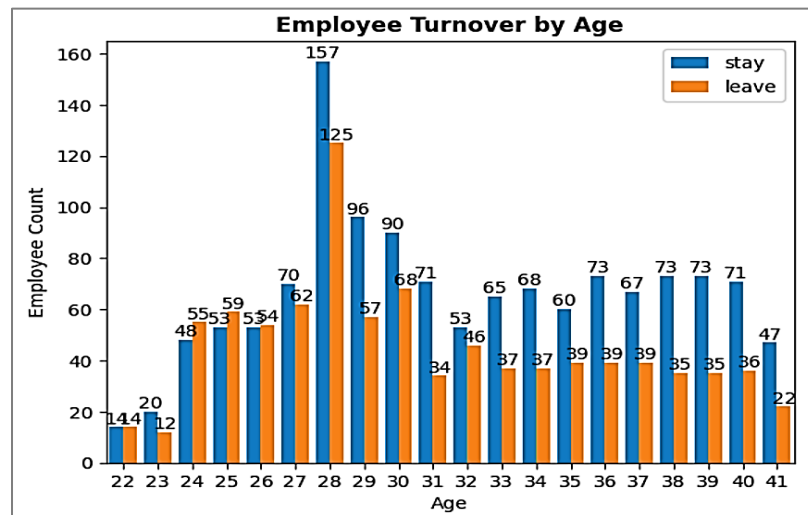


Gambar 4. 3 Visualisasi Atribut *Payment Tier*

Gambar 4.8 adalah visualisasi dari atribut *Payment Tier* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 133 karyawan yang tinggal dan 71 karyawan yang meninggalkan perusahaan dengan tingkat gaji "1", 188 karyawan yang tinggal dan 284 karyawan yang meninggalkan perusahaan dengan tingkat gaji "2" dan 1001 karyawan yang tinggal dan 550 karyawan yang meninggalkan perusahaan dengan tingkat gaji "3".

4. Visualisasi Atribut *Age* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisasi atribut *Age* berdasarkan keputusan karyawan menggunakan *Tools Google Colabulatory*.

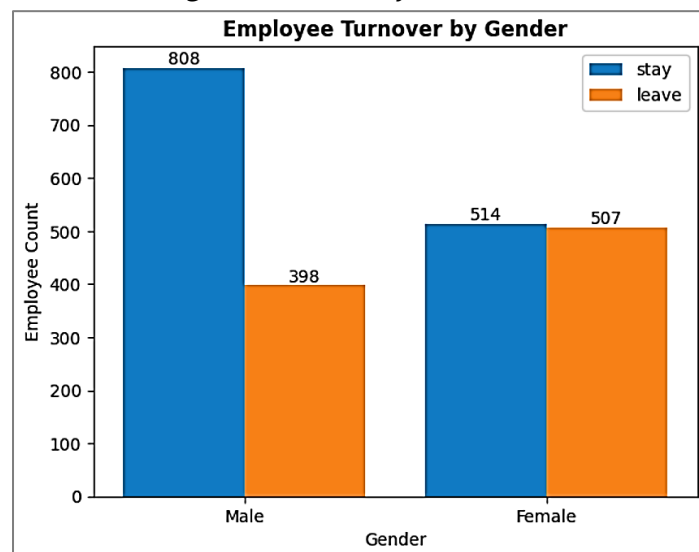


Gambar 4. 4 Visualisasi Atribut Age

Gambar 4.9 menunjukkan jumlah karyawan yang tinggal (*stay*) dan meninggalkan (*leave*) perusahaan berdasarkan kelompok usia. Terlihat bahwa karyawan berusia 28 tahun adalah yang paling banyak, baik yang tinggal maupun yang meninggalkan perusahaan. Secara umum, lebih banyak karyawan yang tetap tinggal dibandingkan dengan yang meninggalkan, terutama di usia-usia yang lebih muda. Namun, seiring bertambahnya usia, perbedaan antara jumlah yang tinggal dan meninggalkan semakin mengecil. Kesimpulannya, usia dapat menjadi faktor penting dalam memahami perilaku turnover karyawan, dengan usia 28 tahun sebagai puncak *turnover*.

5. Visualisasi Atribut *Age* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisai atribut *Age* berdasarkan keputusan karyawan menggunakan *Tools Google Colabatory*.



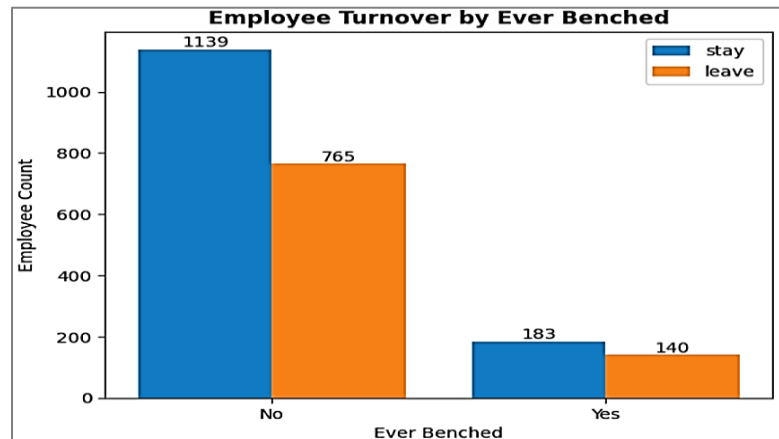
Gambar 4. 5 Visualisasi Atribut Gender

Gambar 4.10 adalah visualisasi dari atribut *Age* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 808 karyawan yang tinggal dan 398

karyawan yang meninggalkan perusahaan dengan jenis kelamin "Male" dan 514 karyawan yang tinggal dan 507 karyawan yang meninggalkan perusahaan dengan jenis kelamin "Female".

6. Visualisasi Atribut *Ever Benched* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisasi atribut *Ever Benched* berdasarkan keputusan karyawan menggunakan *Tools Google Colabatory*.

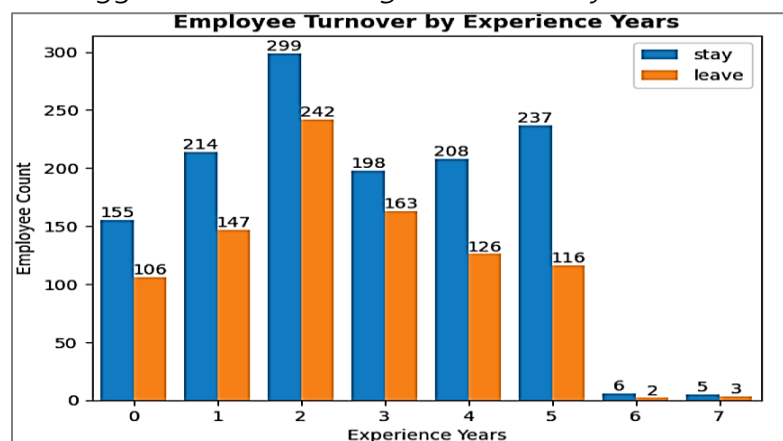


Gambar 4. 6 Visualisasi Atribut Ever Benched

Gambar 4.11 adalah visualisasi dari atribut *Ever Benched* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 1139 karyawan yang tinggal dan 765 karyawan yang meninggalkan perusahaan dengan pernah dicutikan "No" dan 183 karyawan yang tinggal dan 140 karyawan yang meninggalkan perusahaan dengan pernah dicutikan "Yes".

7. Visualisasi Atribut *Experience Years* berdasarkan keputusan karyawan.

Berikut adalah gambar dari bentuk visualisasi atribut *Experience Years* berdasarkan keputusan karyawan menggunakan *Tools Google Colabatory*.



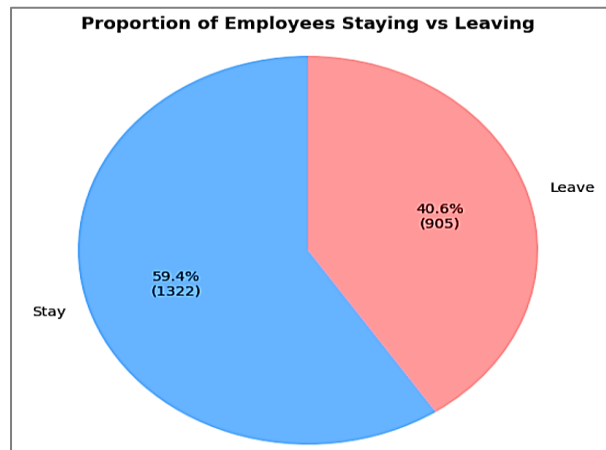
Gambar 4. 7 Visualisasi Atribut Experience Years

Gambar 4.12 menunjukkan jumlah karyawan yang tinggal (stay) dan meninggalkan (leave) perusahaan berdasarkan jumlah tahun pengalaman kerja. Terlihat bahwa karyawan

dengan 2 tahun pengalaman memiliki turnover yang paling tinggi, dengan jumlah yang hampir sama antara yang tinggal dan meninggalkan. Secara umum, turnover cenderung lebih tinggi pada karyawan dengan pengalaman kerja 0 hingga 5 tahun, dan mulai menurun drastis setelah 5 tahun pengalaman. Kesimpulannya, karyawan dengan sedikit pengalaman kerja lebih cenderung meninggalkan perusahaan, sedangkan mereka yang memiliki lebih dari 5 tahun pengalaman cenderung tetap tinggal.

8. Visualisasi Atribut *LeaveOrNot*.

Berikut adalah gambar dari bentuk visualisasi atribut *LeaveOrNot* berdasarkan keputusan karyawan menggunakan *Tools Google Colab*.



Gambar 4. 8 Visualisasi Atribut *LeaveOrNot*

Gambar 4.11 adalah visualisasi dari atribut *Ever Benched* berdasarkan keputusan karyawan. Diketahui bahwa dari 2227 data karyawan terdapat 1322 atau sebesar 59,4% karyawan yang tinggal dan 905 atau sebesar 40,6% karyawan yang meninggalkan perusahaan

1.1. Hasil Klasifikasi *Naïve Bayes* Menggunakan *Phyton*

Pada tahap ini pengujian untuk hitungan akurasi *Naïve Bayes* akan menggunakan bahasa pemrograman *Phyton* dengan *Google Colab* sebagai sarana *cloud* pengkodean untuk menghitung akurasi *Naïve Bayes*. Berikut gambar dari alur coding *phyton* menggunakan google collabs. Berikut gambaran *script phyton* untuk menghitung akurasi *Naïve Bayes*.

Import library

```
[ ] #Mengolah data/membaca file
import pandas as pd
#menyediakan fungsi siap pakai untuk melakukan perhitungan
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

✓ **Import Data**

```
[ ] datadm = pd.read_excel("/content/training Employee.xlsx")
datadm.head()
```

	Education	JoiningYear	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
0	Bachelors	2017	3	34	Male	No	0	0
1	Bachelors	2013	1	28	Female	No	3	1
2	Bachelors	2014	3	38	Female	No	2	0
3	Masters	2016	3	27	Male	No	5	1

Gambar 4. 9 Proses Naïve Bayes Menggunakan Bahasa Pemrograman Phytton

Gambar 4.14 merupakan proses dari *google colaboration* untuk melakukan pengujian akurasi *Naïve Bayes* menggunakan bahasa pemrograman *Phytton*. Berikut adalah hasil klasifikasi dan nilai akurasi *Naïve Bayes* dengan menggunakan *Phytton*.

Hasil Klasifikasi

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.85	0.95	0.90	37
1	0.78	0.54	0.64	13
accuracy			0.84	50
macro avg	0.82	0.74	0.77	50
weighted avg	0.83	0.84	0.83	50

Test Akurasi

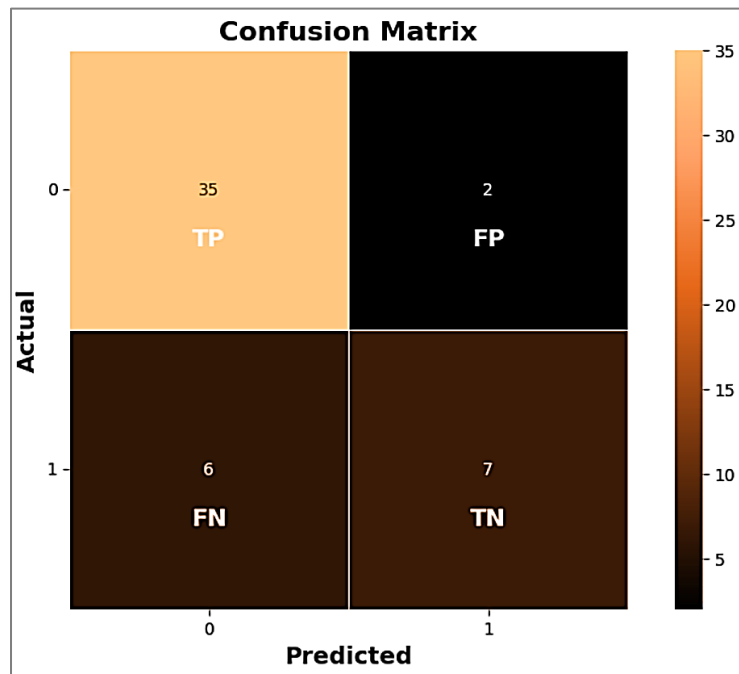
```
[ ] from sklearn import metrics
accuracy = accuracy_score(y_test, y_pred)

print(f"Accuracy: {accuracy * 100:.2f}%")
```

Accuracy: 84.00%

Gambar 4. 10 Hasil Klasifikasi Naïve Bayes Menggunakan Phytton

Gambar 4.15 adalah hasil akurasi klasifikasi algoritma *naïve bayes* dengan menggunakan bahasa pemrograman *phyton*. Mendapatkan hasil akurasi sebesar 84.00%. Berikut adalah gambar confusion matrix klasifikasi algoritma *Naïve Bayes* dengan menggunakan *Phytton*.



Gambar 4. 11 Confusion Matrix

Berikut perhitungan *Confusion Matrix*.

a. Accuracy = $((TP + TN) / (TP + TN + FP + FN)) * 100\%$
 $= ((35 + 7) / (35 + 7 + 2 + 6)) * 100\%$
 $= 84.00\%$

b. Precision = $(TP / (TP + FP)) * 100\%$
 $= (35 / (35 + 2)) * 100\%$
 $= 94.59\%$

c. Recall = $(TP / (TP + FN)) * 100\%$
 $= (35 / (35 + 6)) * 100\%$
 $= 85.37\%$

d. F1-score = $(2 \times (Precision \times Recall)) / (Precision + Recall) * 100\%$
 $= (2 \times (94,59\% \times 85.37\%)) / (94,59\% + 85,37\%) * 100\%$
 $= 89.6\%$

1.2. Hasil Klasifikasi *C4.5* Menggunakan *Phyton*

Pada tahap ini pengujian untuk hitungan akurasi *C4.5* akan menggunakan bahasa pemrograman *Phyton* dengan *Google Colab* sebagai sarana *cloud* pengkodean untuk menghitung akurasi *C4.5*. Berikut gambar dari alur coding *phyton* menggunakan google collabs. Berikut gambaran *script phyton* untuk menghitung akurasi *C4.5*.

```

Import library

[1] #Mengolah data/membaca file
import pandas as pd
#menyediakan fungsi siap pakai untuk melakukan perhitungan
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

Import Data

[2] datadm = pd.read_excel("/content/training Employee.xlsx")
datadm.head()

```

	Education	JoiningYear	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
0	Bachelors	2017	3	34	Male	No		0
1	Bachelors	2013	1	28	Female	No		3
2	Bachelors	2014	3	38	Female	No		2
3	Masters	2016	3	27	Male	No		5

Gambar 4. 12 Proses C4.5 Menggunakan Bahasa Pemrograman Phytion

Gambar 4.17 merupakan proses dari *google colaboration* untuk melakukan pengujian akurasi *Naïve Bayes* menggunakan bahasa pemrograman *Phyton*. Berikut adalah hasil klasifikasi dan nilai akurasi *Naïve Bayes* dengan menggunakan *Phyton*.

```

Hasil Klasifikasi

[38] print(classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	0.90	1.00	0.95	37
1	1.00	0.69	0.82	13
accuracy			0.92	50
macro avg	0.95	0.85	0.88	50
weighted avg	0.93	0.92	0.91	50

```

Test Akurasi

[39] from sklearn import metrics
accuracy = accuracy_score(y_test, y_pred)

print(f"Accuracy: {accuracy * 100:.2f}%")

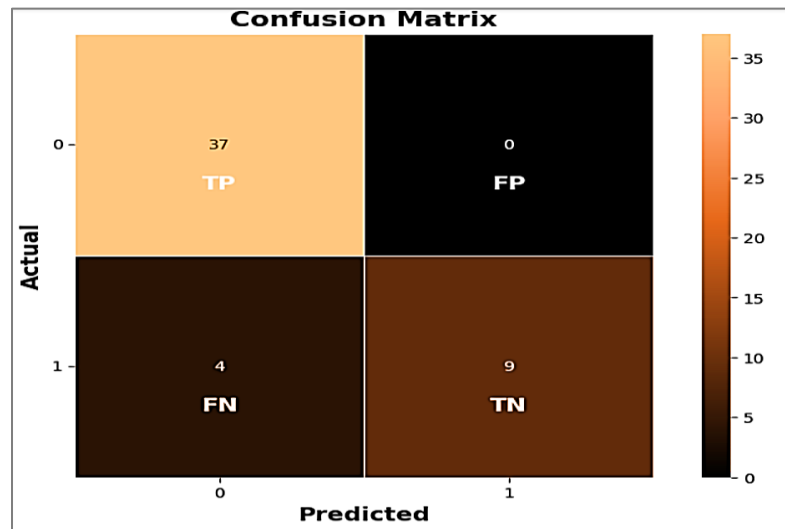
```

Accuracy: 92.00%

Gambar 4. 13 Hasil Klasifikasi C4.5 Menggunakan Phytion

Gambar 4.18 adalah hasil akurasi klasifikasi algoritma *C4.5* dengan menggunakan bahasa pemrograman *phyton*. Mendapatkan hasil akurasi sebesar 92.00%.

Berikut adalah gambar *confusion matrix* klasifikasi algoritma *C4.5* dengan menggunakan *Phyton*.



Gambar 4. 14 Confusion Matrix

Berikut perhitungan *Confusion Matrix*.

a. Accuracy = $((TP + TN) / (TP + TN + FP + FN)) * 100\%$
 $= ((37 + 9) / (37 + 9 + 0 + 4)) * 100\%$
 $= 92.00\%$

b. Precision = $(TP / (TP + FP)) * 100\%$
 $= (37 / (37 + 0)) * 100\%$
 $= 100.00\%$

c. Recall = $(TP / (TP + FN)) * 100\%$
 $= (37 / (37 + 4)) * 100\%$
 $= 90.24\%$

d. F1-score = $(2 \times (Precision \times Recall)) / (Precision + Recall) * 100\%$
 $= (2 \times (100\% \times 90.24\%)) / (100\% + 90.24\%) * 100\%$
 $= 95.84\%$

1.3. Hasil Evaluasi Algoritma *Naïve Bayes* dan *C4.5*

Setelah dilakukan analisis klasifikasi *Naïve Bayes* dan *C4.5* menggunakan bahasa pemrograman *python*. Perbandingan hasil analisis dapat dilihat pada tabel 4.2 .

Tabel 4. 1 Perbandingan Evaluasi

	<i>Naïve Bayes</i>	<i>C4.5</i>
<i>Accuracy</i>	84.00%	92.00%
<i>Precision</i>	94.59%	100.00%
<i>Recall</i>	85.37%	90.24%
<i>F1-score</i>	89.6%	95.84%

Berdasarkan tabel 4.2, dapat disimpulkan bahwa algoritma C4.5 menunjukkan kinerja yang lebih baik dibandingkan dengan Naive Bayes dalam evaluasi model. Hal ini terbukti dengan nilai akurasi, precision, recall, dan F1-score yang lebih tinggi pada C4.5. Algoritma C4.5 mencapai akurasi sebesar 92%, sementara Naive Bayes hanya mencapai 84%. Precision C4.5 bahkan mencapai 100%, sedangkan Naive Bayes hanya 94.59%. Selain itu, nilai recall dan F1-score C4.5 juga lebih tinggi, masing-masing sebesar 90.24% dan 95.84%, dibandingkan dengan Naive Bayes yang memiliki recall 85.37% dan F1-score 89.6%. Hasil ini menunjukkan bahwa C4.5 lebih efektif dalam mengklasifikasikan data pada tugas yang diuji.

SIMPULAN

Berdasarkan pembahasan dapat disimpulkan bahwa 1) Penerapan Algoritma Naïve Bayes dan C4.5: Penelitian berhasil mengimplementasikan algoritma Naïve Bayes dan C4.5 untuk memprediksi keputusan karyawan terkait kemungkinan mereka meninggalkan perusahaan. Data yang digunakan terdiri dari data primer hasil wawancara dengan HRD di CV. Dunia Pustaka dan data sekunder yang diunduh dari Kaggle, yang memungkinkan perbandingan antara kedua algoritma pada data yang berbeda sumbernya. 2) Perbandingan Kinerja Model: Evaluasi model menunjukkan perbandingan kinerja antara Naïve Bayes dan C4.5 berdasarkan metrik akurasi, precision, recall, dan F1-score. Kedua algoritma diuji dengan Python, memberikan gambaran objektif tentang efektivitas masing-masing dalam memprediksi keputusan karyawan secara akurat dan seimbang.

DAFTAR PUSTAKA

- Amir, S. (2018) 'Data Mining Prediksi Kebutuhan Nasabah Dengan Teknik Data Mining Dalam Mendukung Strategi Pemasaran Bank', *Jurnal Bangkit Indonesia*, 7(2), P. 40.
- Andini, R. And Hardiyanti, D.Y. (2024) 'Klasifikasi Karakteristik Turnover Karyawan Menggunakan Algoritma C4. 5', *Jurnal Komtika (Komputasi Dan Informatika)*, 8(2), Pp. 148–161.
- Fadel, M. *Et Al.* (2024) 'Application Of Ensemble Method For Employee Turnover Predictions In Financial Services Company', *Jurnal Teknik Informatika (Jutif)*, 5(3), Pp. 767–775.
- Handayani, S. (2017) 'Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Ginjal', *Jurnal Sistem Informasi*, 6(1), Pp. 34–40.
- Humairosi, L. (2022) 'Implementasi Metode Random Forest Pada Klasifikasi Churn Customer'.

- Maahiroh, L.W. (2024) 'Klasifikasi Turnover Karyawan Menggunakan Algoritma Xgboost (Studi Kasus: Divisi Engineering, Perusahaan Jasa Pertambangan)'. Universitas Islam Indonesia.
- Putri, J.A. (2024) 'Person-Job Fit Menggunakan Decision Trees Algorithm (Studi Kasus Pt Pnr Perusahaan Jasa Pertambangan Pada Divisi Pl)'. Universitas Islam Indonesia.
- Ramadhani, N.H. (2023) 'Perbandingan Algoritma Naïve Bayes Dan Algoritma C4. 5 Dalam Prediksi Kelulusan Mahasiswa Pendidikan Teknik Informatika Dan Komputer Universitas Negeri Jakarta'. Universitas Negeri Jakarta.
- Rianti, R. And Andarsyah, R. (2024) 'Memprediksi Tingkat Atrisi Karyawan Menggunakan Machine Learning', *Jurnal Tekno Insentif*, 18(1), Pp. 39–52.
- Stephen, S. (2022) 'Implementasi Metode Algoritma Naive Bayes Untuk Memprediksi Customer Churn Pada Bisnis Usaha Mikro Kecil Dan Menengah'. Kodeuniversitas041060# Universitasbuddhidharma.
- Tulus, T.H.L., Hadiana, A.I. And Santikarama, I. (2022) 'Sistem Prediksi Awal Terhadap Atrisi Karyawan Menggunakan Algoritma C4. 5', *Informatics And Digital Expert (Index)*, 4(1), Pp. 18–24.
- Utami, S.P. (No Date) 'Klasifikasi Kesehatan Mental Usia Remaja Menggunakan Algoritma Decision Tree Dan Naïve Bayes'. Fakultas Sains Dan Teknologi Uin Syarif Hidayatullah Jakarta.
- Zees, M.E.M. (2023) 'Implementasi Autogluon Dalam Efisiensi Model Prediktif Machine Learning Pada Dataset International Business Machines (Ibm) Human Resource (Hr) Analytics Employee Attrition'. Universitas Islam Indonesia.