



JITE (Journal of Informatics and Telecommunication Engineering)

Available online <http://ojs.uma.ac.id/index.php/jite> DOI : 10.31289/jite.v5i1.5357

Received: 08 June 2021

Accepted: 29 June 2021

Published: 16 July 2021

An Analysis of Contract Employee Performance Assessment Using Machine Learning

Mangapul Siahaan ^{1)*}

1)Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Internasional Batam, Indonesia

*Corresponding Email: mangapul.siahaan@uib.ac.id

Abstrak

Sumber daya manusia menjadi salah satu perhatian utama bagi pengelola semua jenis bisnis baik perusahaan negara maupun perusahaan swasta. Organisasi usaha sangat tertarik dalam menyusun rencana dan eksekusi untuk mempertahankan karyawan kontrak yang berkinerja baik. Dilain pihak banyak nya perpanjangan karyawan kontrak berdasarkan hubungan kedekatan keluarga maupun kedekatan kerabat. Untuk mendapatkan pekerja yang berkinerja baik, dibutuhkan aplikasi *assesment* kinerja karyawan kontrak. Dengan menggunakan pendekatan algoritma machine learning seperti Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes, dan Logistic Regression. Penelitian ini mencoba mencari algoritma yang tepat untuk aplikasi *assesment* kinerja karyawan kontrak dengan nilai input *skill, attitude, responsibility, absent* dan membandingkan nilai akurasi, presisi, recall dan F1 score. Berdasarkan penelitian, model algoritma Random Forest yang terbaik digunakan dalam mengklasifikasi dan prediksi *asement* kinerja karyawan kontrak dengan mencapai nilai akurasi 90.62 %, presisi 72.22 %, recall 75% dan nilai F1 score 71.11 %.

Kata Kunci: Assessment Contract Employee, Machine Learning, Random Forest, Decision Tree, K-Nearest Neighbors.

Abstract

Human resources are one of the main concerns for managers of all types of businesses, both state companies, and private companies. Business organizations are very interested in planning and executing to retain top-performing contract employees. On the other hand, many contract employee extensions are based on close family relationships or close relatives. To get workers who perform well, a contract employee performance assessment application is needed. By using a machine-learning algorithm approach such as Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes, and Logistic Regression. This study tries to find the right algorithm for the application of contract employee performance assessment with the value of input *skill, attitude, responsibility, absence* and compares the values of accuracy, precision, recall, and F1 score. Based on the research, the best Random Forest algorithm model is used in classifying and predicting the performance assessment of contract employees by achieving an accuracy value of 90.62%, precision 72.22%, recall 75%, and F1 score of 71.11%.

Keywords: Assessment Contract Employee, Machine Learning, Random Forest, Decision Tree, K-Nearest Neighbors.

How to Cite: Siahaan, M. (2021). An analysis of Contract Employee Performance Assessment Using Machine Learning. *JITE (Journal Of Informatics And Telecommunication Engineering)*. 5 (1): 122-131

I. PENDAHULUAN

Sumber daya manusia menjadi salah satu perhatian utama bagi pengelola semua jenis bisnis baik perusahaan negara maupun perusahaan swasta. Organisasi usaha sangat tertarik dalam menyusun rencana dan eksekusi untuk mempertahankan karyawan kontrak yang berkinerja baik. Pegawai kontrak adalah pegawai yang bekerja pada organisasi bisnis tertentu dalam jangka waktu yang didasari kesepakatan perjanjian kerja atau Perjanjian Kerja Waktu Tertentu yang disingkat dengan PKWT.

Perjanjian kerja saat ini jangka waktu paling lama 5 tahun seperti yang tertulis dalam Undang-undang Nomor 11 tahun 2020. Dalam mempertahankan masa kerjanya dan dipakai perusahaan, pegawai kontrak harus mempertahankan prestasi kerjanya. Jika penilaian kerja karyawan atau pegawai kontrak memuaskan, maka karyawan kontrak dapat dipertahankan oleh perusahaan untuk berkerja, tetapi jika penilaian kerja karyawan kontrak kurang memuaskan maka perusahaan akan mempertimbangkan mempekerjakan karyawan kontrak tersebut. Evaluasi kerja atau *assessment* merupakan suatu kegiatan untuk mengukur kemampuan pekerja dalam melakukan tugas dan tanggung jawab (Nurchayani & Adnyani, 2016). Salah satu yang menjadi parameter dalam evaluasi kerja karyawan adalah skill atau tingkat kemampuan kerja karyawan, perilaku pekerja atau attitude, tanggung jawab dalam bekerja atau responsibility, dan tingkat kehadiran pekerja atau absen (Yulianto, etc, 2014). Hasil evaluasi kerja karyawan digunakan oleh manajemen dalam mempertimbangkan perpanjangan kontrak karyawan.

PT. XYZ merupakan perusahaan *manufacture* di Kota Batam yang mempekerjakan karyawan kontrak atau PKWT (perjanjian kerja waktu tertentu). Dalam mendapatkan pekerja atau karyawan kontrak yang baik, perusahaan melakukan evaluasi kerja karyawan kontrak pada saat sebelum berakhirnya masa kerja kontrak karyawan. Ada pun yang menjadi parameter dalam evaluasi kerja kontrak karyawan di PT. XYZ adalah *skill, attitude, responsibility, safety* dan tingkat absensi karyawan kontrak. *Assessment* atau evaluasi kerja dilakukan oleh *supervisor* dan *manajer* memutuskan untuk memperpanjang kontrak kerja karyawan berdasarkan prestasi kerja dan volum kerja yang sedang berjalan saat ini. Kendala yang dihadapi adalah, masih terdapatnya perpanjangan kontrak tidak berdasarkan profesional kerja atau hanya berdasarkan kedekatan dengan pimpinan, performansi kerja yang jelek, tingkat absen yang tinggi masih saja dipertahankan pimpinan perusahaan untuk bekerja, begitu sebaliknya karyawan kontrak yang berkinerja bagus tidak diperpanjang oleh pimpinan, dimana seharusnya karyawan yang memiliki tingkat kehadiran yang bagus, performansi kerja yang bagus dan disiplin kerja yang baik, tetapi diputuskan kontrak kerja.

Aplikasi perpanjangan kontrak karyawan harus memiliki kemampuan untuk menganalisa apakah karyawan yang bekerja di PT. XYZ berhak di perpanjang atau tidak sehingga tidak terjadi kecurangan dalam memperpanjang kontrak kerja yang berdasarkan kedekatan pribadi dengan pimpinan. Kemampuan aplikasi sistem informasi "*assessment* karyawan" dalam menganalisa berhaknyanya karyawan kontrak diperpanjang atau tidak dapat diterapkan dengan mengimplementasikan model algoritma machine learning. Tujuan penelitian ini untuk memilih model algoritma *machine learning* yang terbaik seperti algoritma Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes, dan Logistic Regression dengan melihat nilai akurasi, presisi, recall dan F1 score dalam memprediksi penilaian kinerja karyawan apakah karyawan diperpanjang kontrak atau tidak. Berdasarkan kendala diatas, maka penulis melakukan sebuah penelitian mengenai "Analisis Penilaian Kinerja Karyawan Kontrak menggunakan Machine Learning".

II. STUDI PUSTAKA

A. *Machine Learning*

Machine learning adalah teknik dimana mesin dilatih sedemikian rupa sehingga memperoleh kemampuan untuk merespon masukan atau skenario tertentu berdasarkan masukan sebelumnya yang sudah dipelajari (Olalekan, Philip & Martins, 2021). Cukup dengan memberikan sebuah komputer belajar dengan menggunakan teknik statistik. Machine learning membantu komputer untuk bertindak tanpa diprogram secara eksplisit. Tujuan ini untuk mengurangi campur tangan manusia dalam masalah dan skenario yang dapat diandalkan dan membantu menyelesaikan tugas dan masalah yang sangat kompleks dengan sangat mudah dan tanpa melibatkan banyak tenaga manusia. Dengan *Machine Learning* dapat mendeteksi pola secara otomatis (Shwartz & David, 2014). Contoh aplikasi *machine learning* yang telah diterapkan seperti klasifikasi, prediksi, pengenalan gambar, diagnosis medis, pembuatan algoritma, mobil tanpa pengemudi dan lain-lain.

B. *Naives Bayes*

Naive Bayes merupakan teori probabilitas pada kejadian hipotesis dimana tergantung terhadap kejadian lain (Moriesta & Ibrahim, 2017). Metode tersebut ditemukan oleh Thomas Bayes. Teori naive bayes bertujuan untuk mengklasifikasi dimana hampir sama dengan metode neural network dan decision tree, model algoritma naive bayes dapat mencapai nilai akurasi dan kecepatan yang tinggi jika diaplikasikan ke database dengan volume data yang besar (Kusrini dan Luthfi, 2013). Naive Bayes adalah

model algoritma dalam mengklasifikasi secara *probability* atau peluang, menghitung *probability* dengan menjumlah nilai frekuensi dan kombinasi nilai dataset. Persamaan teori naive bayes dapat didefinisikan kedalam rumus (Saleh & Nasari, 2018)

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}, \quad (1)$$

Dimana:

- X : Data class, nilai nya belum diketahui
- H : Hipotesa yang merupakan nilai klas
- P(H\X) : Nilai peluang hipotesa H berdasarkan nilai X
- P(H) : Nilai peluang dari hipotesa H
- P(X\H) : Nilai peluang hipotesis X berdasarkan nilai hipotesa H
- P(X) : Nilai peluang X

C. Logistik Regression

Secara statistik, Logistik Regresion merupakan pendekatan linear dalam memodelkan antara respons skalar (variabel dependen) dengan variabel penjelas (independen). variabel independen disebut regresi linear sederhana, sedangkan variabel indepenen yang lebih dari satu variabel disebut regresi linear berganda. Pada regresi linear, relasi tersebut dimodelkan dalam bentuk fungsi prediktor linear, yang mana parameter modelnya tidak diketahui. Model seperti ini didefinisikan sebagai model linear (Mohri, Rostamizadeh & Talwalkar, 2021).

Model analisis regresi, regresi linear berfokus kepada distribusi *probability* bersyarat dengan respon yang diberikan berupa nilai-nilai prediksi. Distribusi probabilitas merupakan gabungan dari semua variabel yang domain dari analisis multivariate. Teori Logistik Regression dapat dituliskan dengan persamaan::

$$Y = \beta x + \epsilon \quad (2)$$

Dimana:

- y : Vector dari nilai yang diamati
y_i (i=1,2,3,..,n)
- β : Vector parameter dimensi (p+1)
- x : Matriks baris vector x_i (i=1,2,..,n)
- ε : Vector nilai dari ε_i

D. Decision Tree

Decision Tree adalah model klasifikasi yang terlihat seperti pohon dalam representasi grafis dan model tersebut digunakan untuk memprediksi nilai atribut target yang dibangun berdasarkan jumlah entri fitur dari kumpulan. Variabel predictor didefinisikan sebagai internal node yang digunakan sebagai pemisah, setiap leaf node berupa kelas hasil klasifikasi (Han & Kamber, 2012).

E. K-Nearest Neighbor

K-NN adalah algoritma berbasis pembelajaran di mana pelatihan dataset disimpan. Jadi klasifikasi record baru yang tidak terklasifikasi diperoleh dengan membandingkan record yang paling mirip dengan training set. Selain digunakan untuk klasifikasi, algoritma K-NN juga digunakan untuk estimasi dan prediksi. Klasifikasi K-NN berdasarkan kepada kedekatan lokasi data dengan data lain (Prasetyo, 2012). Persamaan teori K-NN dapat didefinisikan kedalam bentuk rumus persamaan.

$$d(x_i, x_j) = \left(\sum_{k=1}^K |x_{ik} - x_{jk}|^r \right)^{1/r} \quad (3)$$

Dengan

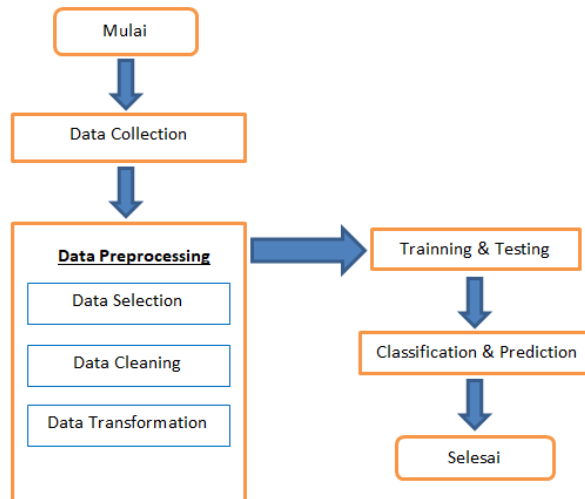
- X_{ik} : Data test ke i di variabel ke -k
- X_{jk} : Data training j di variabel ke-k
- D(X_i, X_j) : Jarak
- K : Dimensi data variabel bebas

F. Random Forest

Random forest merupakan model klasifikasi supervised learning dimana banyak digunakan untuk regresi dan klasifikasi. Umumnya digunakan untuk memprediksi atau mengklasifikasi. Random Forest yang terdiri dari banyak tree dapat membuat keputusan bercabang. Random forest sendiri merupakan salah satu metode Decision Tree.

III. METODE PENELITIAN

Tahapan penelitian dirangkai secara sistematis dimulai dari pengumpulan data, data preprocessing, training dan testing, dan mengklasifikasi dan melakukan prediksi.



Gambar 1. Tahapan Penelitian

A. Data Collection

Pengumpulan data diambil dari aplikasi internal PT. XYZ dimana aplikasi melakukan *asessment performance* karyawan kontrak yang dilakukan oleh *supervisor*. *Supervisor* melakukan assesment terhadap karyawan kontrak dengan *parameter skill, attitude, responsibility, safety dan absent*. Data absen diambil dari data finger karyawan. Data diambil dari database PostgreSQL, adapun data yang dikumpul berjumlah 156 data. Parameter dataset ini yaitu: *nik, gender, n_marital, n_education, grade_code, departemen, jabatan, join_date, end_contract, workingincompany, rating_skill, rating_attidute, rating_responsibility, rating_safety, rating_absen* dan *performance*.

Detail IOM 009 / Detail Assesment IOM 0023/HR/CTE/V/2021

| Name | Skill (40%) | Attitude (15%) | Responsibility (15%) | Safety (10%) | Absent (20%) | Total | Note From HR |
|---------------------|-------------|----------------|----------------------|--------------|--------------|-------|--------------|
| ADEN REANDY PERWIRA | ★★★★ | ★★★ | ★★★ | ★★★★ | ★★ | 3.2 | |
| DIDIT ADITAMA | ★★★ | ★★★ | ★★★ | ★★★★ | * | 2.6 | |
| SAHLAN | ★★★ | ★★★ | ★★ | ★★★★ | ★★★★ | 3.05 | |
| SUZALI | ★★★ | ★★★ | ★★ | ★★ | ★★★★ | 2.95 | |

Please advise whether the contract will be continued or discontinued. If will be continued, please fill in "Employee Request Form" attached.

Gambar 2. Aplikasi assesment karyawan kontrak

B. Data Preprocessing

Data preprocessing melakukan perbaikan data. Data yang bernilai kosong dicek terlebih dahulu, dan kolom yang tidak digunakan untuk proses data dihapus. Tujuannya untuk mendapatkan hasil prediksi yang akurat. Semua kolom dipastikan sudah dalam bentuk numerik dengan mentransformasikan data kolom yang object ke numerik.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---------|----------|-----------|-----------|-----------|----------|------------|------------|--------------|-----------|------------|------------|------------|------------|------------|-------------|
| nik | n_gender | n_marital | n_educati | grade_coc | departme | jabatan | join_date | end_contract | workingin | rating_ski | rating_att | rating_res | rating_saf | rating_abs | performance |
| 2103971 | Male | Married | SLTA | 2 | FABRICAT | FORKLIFT | 3/8/2021 | 6/7/2021 | 3 | 5 | 5 | 5 | 5 | 3 | 4.06 |
| 2102967 | Male | Married | SMK | 2 | FABRICAT | FITTER STF | 2/18/2021 | 5/17/2021 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2102966 | Male | Married | SMK | 2 | FABRICAT | FITTER STF | 2/18/2021 | 5/17/2021 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2102965 | Male | Married | SMK | 2 | FABRICAT | FITTER STF | 2/18/2021 | 5/17/2021 | 3 | 5 | 5 | 5 | 5 | 4 | 4.08 |
| 2102961 | Male | Married | SMK | 2 | FABRICAT | WELDER G | 2/1/2021 | 5/15/2021 | 3 | 4 | 4 | 4 | 4 | 4 | 4 |
| 2101956 | Male | Married | SMK | 2 | FABRICAT | WELDER S | 1/20/2021 | 5/19/2021 | 4 | 3 | 3 | 3 | 3 | 1 | 2.06 |
| 2101955 | Male | Married | SMK | 2 | FABRICAT | WELDER S | 1/18/2021 | 5/17/2021 | 4 | 4 | 3 | 3 | 4 | 3 | 3.05 |
| 2101949 | Male | Married | SMK | 3 | QUALITY | C WELDING | 1/6/2021 | 6/5/2021 | 5 | 3 | 3 | 4 | 4 | 3 | 3.25 |
| 2101947 | Male | Married | SLTA | 2 | FABRICAT | BLASTER | 1/4/2021 | 5/31/2021 | 5 | 3 | 3 | 3 | 3 | 4 | 3.02 |
| 2012943 | Male | Divorce | Strata 1 | 4 | FABRICAT | CONSTRU | 12/28/2020 | 5/27/2021 | 5 | 4 | 4 | 3 | 4 | 1 | 3.25 |
| 2012940 | Male | Married | SLTA | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 4 | 4 | 4 | 4 | 4 | 4 |
| 2012939 | Male | Married | SMK | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 3 | 4 | 4 | 4 | 4 | 3.06 |
| 2012938 | Male | Single | SMK | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 3 | 4 | 4 | 4 | 5 | 3.08 |
| 2012937 | Male | Married | SLTA | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 3 | 4 | 4 | 4 | 5 | 3.08 |
| 2012936 | Male | Single | SMK | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 3 | 4 | 4 | 4 | 5 | 3.08 |
| 2012935 | Male | Married | SMK | 2 | FABRICAT | GRINDER | 12/21/2020 | 5/20/2021 | 5 | 3 | 4 | 4 | 4 | 4 | 3.06 |
| 2012934 | Male | Single | SLTA | 2 | FABRICAT | PAINTER | 12/19/2020 | 5/31/2021 | 5 | 4 | 4 | 4 | 4 | 4 | 4 |
| 2012932 | Male | Married | SLTA | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 5 | 5 | 5 | 5 | 4 | 4.08 |
| 2012931 | Male | Married | SMK | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 4 | 5 | 5 | 4 | 5 | 4.05 |
| 2012930 | Male | Married | SLTA | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 1 | 1 | 1 | 1 | 4 | 1.06 |
| 2012929 | Male | Single | SLTA | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2012928 | Male | Married | SMK | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 4 | 5 | 5 | 4 | 4 | 4.03 |
| 2012927 | Male | Married | SMK | 2 | FABRICAT | BLASTER | 12/19/2020 | 5/31/2021 | 5 | 5 | 5 | 5 | 5 | 4 | 4.08 |
| 2012926 | Male | Single | SMK | 2 | FABRICAT | GRINDER | 12/17/2020 | 5/15/2021 | 5 | 3 | 4 | 4 | 4 | 5 | 3.08 |

Gambar 3. Data Preprocessing

C. Training & Testing

Melakukan training data dan testing data untuk membangun model. Dataset tersebut di split dengan data train berjumlah 80% dan data test sebesar 20%.

D. Klasifikasi dan Prediksi

Memilih model yang tepat dalam klasifikasi dan prediksi dan membandingkan hasil mana yang lebih akurat dari hasil model algoritma yang digunakan. Model algoritma klasifikasi dan prediksi yang digunakan adalah Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes, dan Logistic Regression.

IV. HASIL DAN PEMBAHASAN

A. Data Preprocessing

Dataset yang akan digunakan pada penelitian ini adalah dataset yang diunduh dari aplikasi internal yaitu aplikasi sistem informasi *assessment* performansi karyawan kontrak PT. XYZ. Data berjumlah 156 baris dan terdiri dari 16 kolom yaitu *nik*, *gender*, *n_marital*, *n_education*, *grade_code*, *departemen*, *jabatan*, *join_date*, *end_contract*, *workingincompany*, *rating_skill*, *rating_attidute*, *rating_responsibility*, *rating_safety*, *rating_absen* dan *performance*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 16 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   nik                                   156 non-null    int64
1   n_gender                             156 non-null    object
2   n_marital                             156 non-null    object
3   n_education                           155 non-null    object
4   grade_code                            156 non-null    int64
5   department                            156 non-null    object
6   jabatan                               156 non-null    object
7   join_date                             156 non-null    object
8   end_contract                          156 non-null    object
9   workingincompany                      156 non-null    int64
10  rating_skill                           156 non-null    int64
11  rating_attitude                        156 non-null    int64
12  rating_responsibility                  156 non-null    int64
13  rating_safety                          156 non-null    int64
14  rating_absent                          156 non-null    int64
15  performance                            156 non-null    float64
dtypes: float64(1), int64(8), object(7)
memory usage: 19.6+ KB
```

Gambar 4. Dataset data penelitian.

Data Selection

Data yang digunakan yaitu *nik, gender, n_marital, n_education, grade_code, departemen, workingincompany, rating_skill, rating_attitude, rating_responsibility, rating_safety, rating_absen*. Sedangkan untuk target, data yang digunakan adalah *performance*.

Data Cleaning

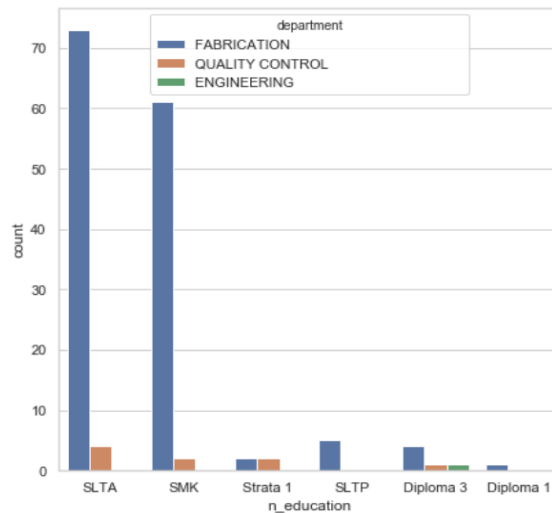
Dataset yang didownload dicek terlebih dahulu apakah ada data kosong. Didapat data education memiliki data kosong, sehingga pada baris yang kosong untuk kolom education diisi dengan education SLTA.

Data transformation

Dataset yang bertipe object harus di transformasikan ke numerik agar perhitungan nilai klasifikasi dan prediksi menghasilkan nilai yang akurat. Adapun data objek yang perlu di transformasi yaitu: *n_gender, n_marital* dan *grade_code* sehingga perlu dilakukan transformasi dan menghasilkan nilai numerik 1,2 dan 3 sesuai dengan jenis dari data tersebut. Sedangkan untuk jabatan, *join_date*, dan *end_contract* dilakukan penghapusan kolom.

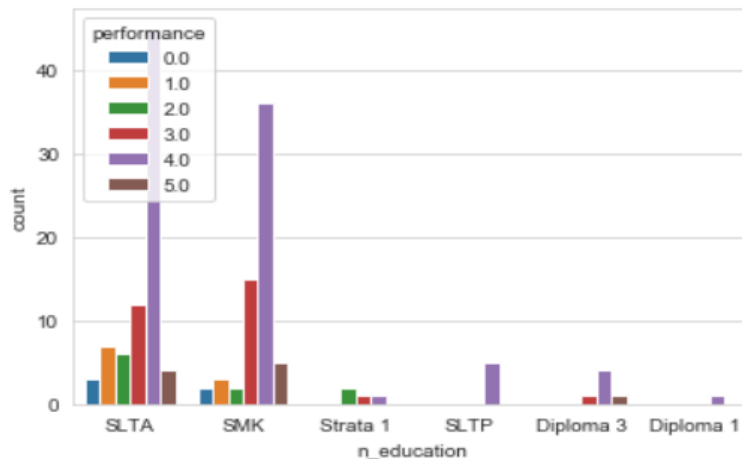
B. Visualisasi data

Perbandingan departemen dengan tingkat Pendidikan, didepartemen fabrikasi lebih dominan berpendidikan SLTA dan SMK.



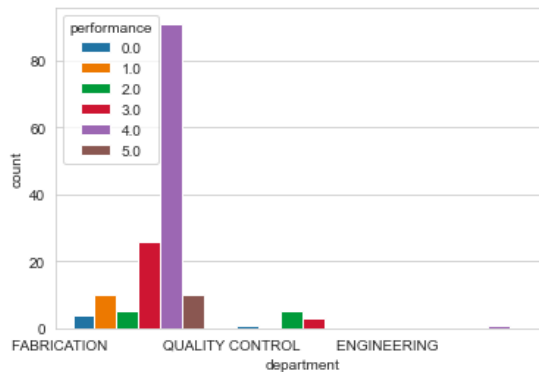
Gambar 5. Data visualisasi tingkat Pendidikan dan departemen

Perbandingan tingkat Pendidikan dengan performansi kinerja karyawan, umumnya tingkat Pendidikan SLTA lebih baik performansinya kinerja karyawan dibandingkan tingkat Pendidikan SMK dan diploma.



Gambar 6. Data visualisasi departemen dan performansi karyawan.

perbandingan departemen dengan nilai performasi karyawan kontrak. Departemen Quality Control memiliki nilai assesment rata-rata 4, dimana nilai assessment performansi lebih baik dibandingkan departemen lainnya.



Gambar 7. Data visualisasi departemen dan performansi karyawan.

C. Training & Testing

Dalam membangun model, dilakukan split data antara data training dan testing data. Dalam penelitian ini dataset displit menjadi data train sebesar 80% dan data test sebesar 20%.

D. Klasifikasi dan Prediksi

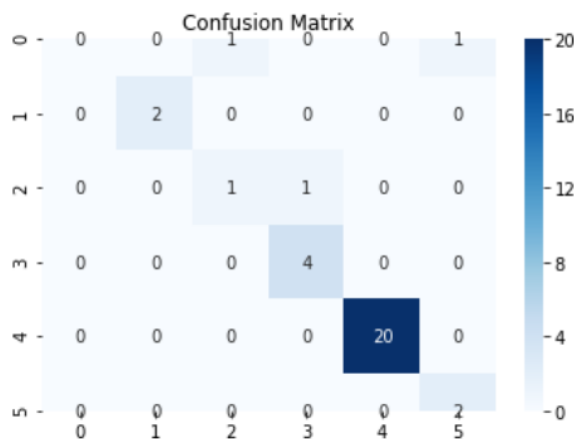
Pada tahap ini melakukan klasifikasi dan prediksi dengan menggunakan model algoritma Random Forest, Decision Tree, K-Nearest Neighbors, Naïve Bayes, dan Logistic Regression. Tujuan nya adalah melakukan perbandingan nilai akurasi, nilai presisi, nilai recall dan nilai f1 score dari masing-masing model algoritma. Semakin tinggi nilai akurasi maka algoritma yang digunakan semakin baik (Mittal & Gill, 2014).

1. Decision Tree

klasifikasi dataset dengan menggunakan model Decision Tree. Dengan menggunakan model algoritma Decision Tree, didapat hasil skor

Tabel 1. Hasil perhitungan Decision Tree

| | Algoritma Decision Tree |
|-----------|-------------------------|
| Accuracy | 84.38% |
| Precision | 66.11% |
| Recall | 73.33 % |
| F1 | 68.94 % |



Gambar 8. Confusion Matrix.

Berdasarkan gambar diatas dapat diketahui nilai total true positif (TP) sebanyak 27 dan nilai false positif (FP) sebanyak 5, sehingga bisa menghitung nilai akurasi sebagai berikut

$$\begin{aligned} \textit{Precision} &= TP / (TP + FP) \\ &= 27 / (27+5) \\ &= 0.84 \end{aligned}$$

$$\begin{aligned} \textit{Recall} &= TP / (TP + FN) \\ &= 27 / (27+5) \\ &= 0.84 \end{aligned}$$

$$\textit{Accuracy} = \textit{Precision} = \textit{Recall} = 0.84$$

2. Naïve Bayes

Hasil klasifikasi dataset dengan menggunakan model Naïve Bayes. Dengan menggunakan algoritma Naïve Bayes, didapat nilai skor

Tabel 2. Hasil perhitungan Naïve Bayes

| | Algoritma Naïve Bayes |
|-----------|-----------------------|
| Accuracy | 78.12 % |
| Precision | 49.44 % |
| Recall | 50 % |
| F1 | 49.34 % |

3. Random Forest

Hasil klasifikasi dataset dengan menggunakan model Random Forest. Dengan menggunakan algoritma Random Forest, didapat nilai skor

Tabel 3. Hasil perhitungan Random Forest

| | Algoritma Random Forest |
|-----------|-------------------------|
| Accuracy | 90.62 % |
| Precision | 72.22 % |
| Recall | 75 % |
| F1 | 71.11 % |

4. K-Nearest Neighbors

Hasil klasifikasi dataset dengan menggunakan model K-Nearest Neighbors. Dengan menggunakan algoritma K-Nearest Neighbors, didapat nilai skor

Tabel 4. Hasil perhitungan Neighbors

| | Algoritma K-Nearest Neighbors |
|-----------|-------------------------------|
| Accuracy | 56.25 % |
| Precision | 16.03 % |
| Recall | 21.67 % |
| F1 | 17.77 % |

5. Logistik Regression

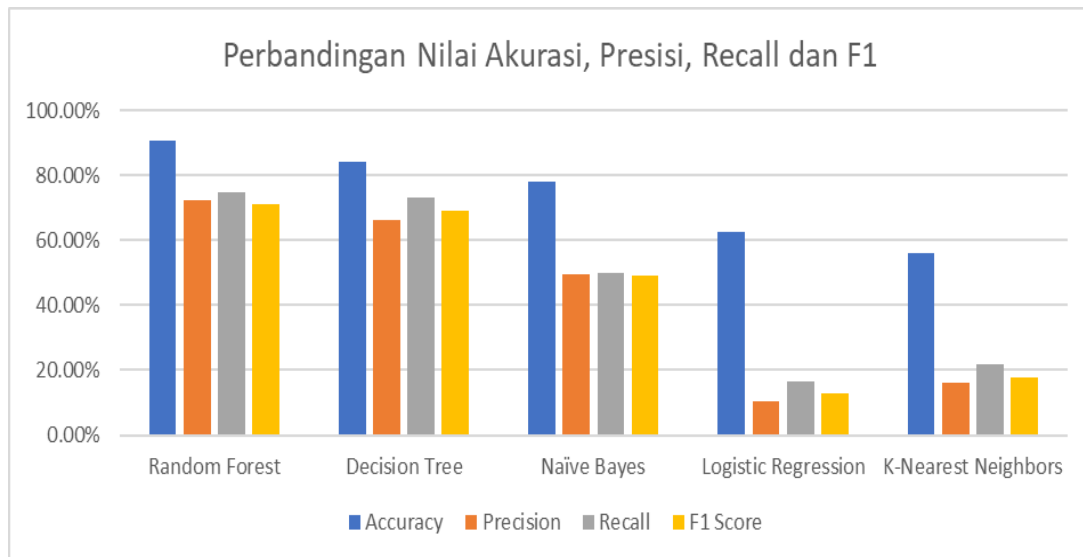
Hasil klasifikasi dataset dengan menggunakan model Logistik Regression. Dengan menggunakan algoritma Logistik Regression, didapat nilai skor

Tabel 5. Hasil perhitungan Logistik Regression

| | Algoritma Logistik Regression |
|-----------|-------------------------------|
| Accuracy | 62.5 % |
| Precision | 10.42 % |
| Recall | 16.67 % |
| F1 | 12.82 % |

Tabel 6. Hasil perbandingan model algoritma

| Klasifikasi | Accuracy | Precision | Recall | F1 Score |
|---------------------|----------|-----------|---------|----------|
| Random Forest | 90.62 % | 72.22 % | 75 % | 71.11 % |
| Decision Tree | 84.38 % | 66.11 % | 73.33 % | 68.94 % |
| Naïve Bayes | 78.12 % | 49.44 % | 50 % | 49.34 % |
| Logistic Regression | 62.50 % | 10.42 % | 16.67 % | 12.82% |
| K-Nearest Neighbors | 56.25 % | 16.03 % | 21.67 % | 17.77 % |



Gambar 9. Grafik perbandingan nilai akurasi, presisi, recall dan F1

V. SIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dengan menggunakan 5 model algoritma machine learning, maka didapat model algoritma K-Nearest Neighbors tidak tepat digunakan untuk aplikasi *assessment* karyawan kontrak karena menghasilkan nilai akurasi dan nilai presisi yang sangat kecil, Untuk algoritma Random Forest lebih tepat diimplementasikan pada aplikasi *assessment* performansi karyawan kontrak dengan nilai akurasi mencapai 90.62%, sedangkan nilai akurasi yang dihasilkan dari *confusion matrix* dari total pengujian yang dilakukan menghasilkan nilai sebesar 84%, terdapat perbedaan signifikan antara nilai akurasi perhitungan dengan total penjumlahan pengujian.

DAFTAR PUSTAKA

- Agrawal, D., & Dahiya, P. (2018). Comparisons of classification algorithms on seeds dataset using machine learning algorithm. *Compusoft*, 7(5), 2760–2765. <https://doi.org/10.6084/ijact.v7i5.720>
- Awujoola, O., Philip O Odion, Martins E Irhebhude and Halima Aminu. (2021). Performance Evaluation of Machine Learning Predictive Analytical Model for Determining the Job Applicants Employment Status, *Malaysian Journal of Applied Sciences*, 6(1), pp. 67-79. doi: 10.37231/myjas.2021.6.1.276.
- E. Prasetyo. (2012). *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*, Yogyakarta: ANDI Yogyakarta.
- Hosmer and Lemeshow. (2018). *Applied Logistic Regression*, USA: John Wiley & Sons
- J. Han, M. Kamber and J. Pei. (2012). *Data Mining Concepts and Techniques*, 3rd ed., USA: Morgan Kaufman.
- Kusrini, Luthfi, Emha Taufiq. (2013). *Algoritma Data Mining*. Yogyakarta: Andi Offset
- M. Mohri, A. Rostamizadeh, dan A. Talwalkar. (2012). *Foundations of Machine Learning (Adaptive Computation and Machine Learning series)*

- Mittal, P., & Gill, N. S. (2014). A Comparative Analysis Of Classification Techniques On Medical Data Sets, *IJRET: International Journal of Research in Engineering and Technology*, 3 (6): 454-460. <https://doi.org/10.15623/ijret.2014.0306085>
- Moriesta, Elsy, Selviani; Ibrahim, Ali. (2017). ANALISIS PENYARINGAN EMAIL SPAM MENGGUNAKAN METODE NAIVE BAYES. *Annual Research Seminar (ARS)*, [S.l.], v. 3, n. 1, p. 45-48. <http://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1677>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Ni Made Nurcahyani and I. G. A. D. Adnyani. (2016). Pengaruh Kompensasi dan Motivasi Terhadap Kinerja Karyawan Dengan Kepuasan Kerja Sebagai Variabel Intervening, *E-jurnal Manajemen.*, vol. 5, no. 1, pp. 500–532.
- Saleh, A., & Nasari, F. (2018). Penggunaan Teknik Unsupervised Discretization pada Metode Naive Bayes dalam Menentukan Jurusan Siswa Madrasah Aliyah. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 353. <https://doi.org/10.25126/jtiik.201853705>
- Shwartz, S. S., & David, S. B. (2014). *Understanding Machine Learning From Theory To Algorithms*. Cambridge University Press
- R. Sarno and J. Sidabutar.(2015). Comparison of Different Neural Network Architectures for Software Cost Estimation, in *International Conference on Computer, Control, Informatics and Its Applications Comparison*, pp. 68–73.
- W. Julianto, R. Yunitarini, and M. K. Sophan.(2014). Algoritma C4.5 Untuk Penilaian Kinerja Karyawan, *Scan*, vol. Vo. IX, no. No. 2, pp. 33–39.