

Classification of Super Air Jet Initial Cabin Crew Candidates Using K-Nearest Neighbor (KNN) Method

¹Ahmad Jurnaidi Wahidin, ² Reza Maulana

¹Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika

²Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri

Email: ¹ahmad.ajn@bsi.ac.id

Abstract

At the time of the corona virus outbreak that hit Indonesia and the world which had an impact on various sectors including transportation, optimistically that the Indonesian domestic flight market was still open with strong demand, a new airline, Super Air Jet (SAJ) was created in March 2021. In an effort to improve SAJ services screening qualified human resources, including the process of screening candidates for initial cabin crew. To support this process, it is necessary to have a method used to classify candidates for initial cabin crew at the administrative stage. K-Nearest Neighbor (KNN) which is one method for classifying is expected to provide a solution to the problems discussed. This study uses 10 training data that have 8 criteria to predict categories in 6 test data. From the calculation results using a value of $k=5$, the same results as the label on the initial test data resulted, which resulted in 2 data with the Stop prediction class and 4 data with the Advanced prediction class. To measure the performance of the KNN method, a test was conducted using a confusion matrix which resulted in a 100% accuracy value, a 100% precision value and a 100% recall value.

Keywords: *K-Nearest Neighbor, Classification, Data Mining, Initial Cabin Crew, Super Air Jet*

1. INTRODUCTION

Wabah Virus Corona atau Covid-19 ditetapkan sebagai Bencana Nasional pada 14 maret 2020 oleh Presiden melalui Kepala Badan Nasional Penanggulangan Bencana Doni Monardo di Gedung BNPB [1]. Wabah covid-19 membawa dampak ke berbagai sektor, termasuk perekonomian dan bisnis, beberapa bidang yang terdampak negatif adalah perdagangan, investasi, pariwisata sampai transportasi.

Disaat Indonesia dan Dunia sedang terjadinya wabah virus covid-19 terdapat maskapai penerbangan terbaru yang mencoba peruntungan di industri penerbangan Indonesia yaitu maskapai Super Air Jet yang disingkat SAJ, didirikan pada maret 2021 dengan mengantongi Sertifikat Operator Penerbangan (*Air Operator Certificate/ AOC*) nomor 121-060 [2]. Direktur Utamanya, Ari Azhari, mengatakan maskapai swasta baru ini sepenuhnya didanai dan dimiliki oleh penduduk setempat. “Maskapai ini merupakan maskapai swasta baru dalam pelayanan angkutan penumpang berjadwal harian yang kepemilikan dan penyertaan modalnya berasal dari masyarakat setempat, tulisnya dalam keterangan tertulis, Senin, 3 Mei 2021”[3]. Azhari mengatakan maskapai ini didirikan berdasarkan optimisme bahwa pasar penerbangan domestik

Indonesia masih terbuka dengan permintaan yang kuat untuk kategori ini, terutama permintaan yang cepat dari kaum milenial[3].

Maskapai dengan *callsign* "IU" dari International Air Transport Association (IATA) dan "SJV" dari International Civil Aviation Organization (ICAO) itu didirikan atas dasar optimis bahwa peluang pasar khususnya kebutuhan penerbangan dalam negeri (domestik Indonesia) masih ada dan terbuka luas[4]. Super Air Jet (SAJ) adalah maskapai penerbangan bertarif rendah yang berbasis di Jakarta, Indonesia. SAJ memberikan pengalaman perjalanan dengan penerbangan berkualitas premium, jaringan konektivitas yang tinggi, terpercaya, dan transportasi dengan harga yang terjangkau untuk generasi modern[5].

Data mining dapat diterapkan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual[6]. Klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah learning (fase training), dimana algoritma klasifikasi dibuat untuk menganalisa data latih lalu direpresentasikan dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data uji digunakan untuk memperkirakan akurasi dari rule klasifikasi [7].

K-Nearest Neighbor (KNN) yang merupakan salah satu metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut[8]. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik *clustering*, yaitu mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga terdekat. Proses metode *K-Nearest Neighbor* (KNN) dilakukan dengan mencari kelompok k objek dalam data uji yang paling dekat dengan objek pada data baru atau data uji. KNN adalah algoritma *supervised learning* yang maksudnya algoritma ini menggunakan data yang telah ada dan outputnya telah diketahui

Implementasi data mining pada maskapai penerbangan pernah dilakukan pada penelitian yang memanfaatkan k-means dalam pengelompokan Pilot pada maskapai Lion Air, dimana penelitian tersebut membagi data kedalam beberapa kelompok dengan kemiripan karakteristik[9].

Sedangkan data mining klasterisasi menggunakan metode *K-Nearest Neighbor* pernah digunakan pada penelitian seleksi pegawai dan dosen di UMRI. Dimana UMRI perlu mengimplementasikan *E-recruitment* sebagai salah satu teknologi yang bisa digunakan untuk menjaring pegawai dan dosen yang profesional. Penelitian ini bertujuan bagaimana penerapan metode klasifikasi *K-Nearest Neighbor* (K-NN) dalam system e-recruitment untuk seleksi awal pegawai dan dosen UMRI. Metode K-NN akan menghitung tingkat kemiripan dengan cara

mengukur jarak antara persyaratan yang ditetapkan oleh bagian kepegawaian UMRI dengan data yang dimiliki oleh calon pegawai dan dosen[10].

Penelitian dengan judul Implementasi Metode *K-Nearest Neighbor (KNN)* Untuk Seleksi Calon Karyawan Baru membahas pengembangan dan pembuatan program seleksi calon karyawan berbasis web yang memudahkan dalam klasifikasi data karyawan baru yang termasuk dalam kategori lolos atau tidak lolos. Dengan menggunakan sistem ini diharapkan dapat membantu pihak HRD dalam mengolah data karyawan dengan tepat dan akurat. Dan hasil penghitungan algoritma K- Nearest Neighbor dengan nilai parameter $K = 7$ menggunakan metode Euclidean Distance didapat nilai akurasi sebesar 91%, nilai presisi sebesar 87%, dan nilai recall sebesar 100%[7].

Pada penelitian ini akan menerapkan metode *K-Nearest Neighbor (KNN)* untuk klasifikasi calon *initial cabin crew* pada proses *recruitment* maskapai Super Air Jet ditahap administrasi atau pemberkasan, dimana tahap ini menyaring pelamar berdasarkan persyaratan yang diberikan oleh tim *recruitment* sebelum masuk ke tahap selanjutnya yaitu tes psikologi.

2. METHODS

Dalam sebuah penelitian diperlukan permodelan agar penelitian lebih terarah. Penelitian ini menerapkan metode knn untuk klasifikasi calon initial cabin crew dengan metodologi penelitian sebagai berikut:

2.1 Pengumpulan Data

Teknik pengumpulan data menggunakan studi literatur, metode studi literatur adalah serangkaian kegiatan yang berkenaan dengan metode pengumpulan data pustaka, membaca dan mencatat, serta mengelolah bahan penelitian[11]. Penelitian ini merupakan penelitian yang bertujuan menjelaskan fenomena yang ada dengan menggunakan angka-angka, dimana untuk melakukan klasifikasi terhadap calon initial cabin crew menggunakan data-data yang diolah menggunakan metode KNN.

2.2 Persiapan Data

Pada tahap ini dimana sudah dilakukan *data cleaning* kemudian menentukan atribut yang selanjutnya disebut kriteria, dan menentukan pembobotan dari setiap kriteria dengan besaran angka 0 sampai 3.

2.3 Inisialisasi Data

Data yang akan melalui proses *preprocessing* ditransformasikan dan disimpan ke dalam bentuk yang akan bisa diterapkan pada penghitungan metode knn, dilakukan inisialisasi pada atribut dan perubahan tipe data menjadi numerik.

2.4 Normalisasi Data

Normalisasi merupakan teknik analisis data yang mengorganisasikan atribut-atribut data dengan cara mengelompokkan sehingga terbentuk entitas yang non-redundant, stabil dan fleksibel. Normalisasi dapat dihitung dengan persamaan (1).

$$Normalisasi = \frac{data_x - data_{min}}{data_{max} - data_{min}} \quad (1)$$

dimana:

- $data_x$ = baris pertama dari satu atribut data latih
- $data_{min}$ = data terkecil dari seluruh data pada atribut data x
- $data_{max}$ = data terbesar dari seluruh data pada atribut data x

2.5 Menghitung Euclidean Distance

Untuk mencari jarak antara dua titik yaitu titik pada data latih dan titik pada data uji, maka digunakan rumus Euclidean Distance dengan persamaan (2), sebagai berikut:

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (2)$$

Dimana:

x_i adalah data yang atributnya sudah di normalisasi dan

y_i adalah data uji baru selain data latih.

Persamaan (2) adalah rumus untuk menghitung jarak pada metode *K-Nearest Neighbor* (*KNN*) dengan *Eucliden Distance*. Dengan melihat konsep persamaan (2), maka harus memiliki 2 jenis data yaitu data latih dan data uji untuk diklasifikasikan jarak *euclidean*nya.

2.6 Menentukan Tetangga Terdekat Berdasarkan Jarak Minimum ke K

Tahapan ini adalah menentukan atau pemilihan nilai *distance* terkecil yang masuk peringkat sejumlah nilai k . Misalnya jika nilai k adalah 5 maka harus mencari nilai *distance* terkecil sebanyak lima nilai. Kemudian menentukan kategori/label pada kelima nilai yang sudah diketahui.

2.7 Menetapkan Kelas

Menjumlahkan kategori/label yang sudah didapat pada tahap sebelumnya, kemudian menggunakan kategori/label mayoritas dari tetangga terdekat berdasarkan nilai k sebagai nilai prediksi data uji

2.8 Pengujian

Pada penelitian pengujian dilakukan dengan *Confusion Matrix*, *Confusion Matrix* adalah suatu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.

Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu True Positif, True Negatif, False Positif, dan False Negatif [12]. Contoh *Confusion Matrix* ditampilkan pada gambar 1.

| | | Actual Values | |
|------------------|--------------|---------------|--------------|
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Gambar 1 *Confusion Matrix*

Akurasi menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar, akurasi dapat dihitung dengan persamaan (3).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

Presisi menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model, presisi dapat dihitung menggunakan persamaan (4)

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \quad (4)$$

Recall atau *sensitivity*: menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi, penghitungan recall menggunakan persamaan (5)

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \quad (5)$$

dimana:

TP: *True positives*, merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.

TN: *True negatives*, merupakan jumlah data dengan kelas negatif yang diklasifikasikan negatif.

FP: *False positives*, merupakan jumlah data dengan kelas positif diklasifikasikan negatif.

FN: *False negatives*, merupakan jumlah data dengan kelas negatif diklasifikasikan positif.

3. RESULTS AND DISCUSSION

Penelitian ini bertujuan untuk mengklasifikasikan calon initial cabin crew menggunakan metode *k-nearest neighbor* (knn) dengan menggunakan nilai k=5, nilai k adalah jumlah tetangga paling dekat. Adapun Langkah-langkah penelitian sebagai berikut:

3.1. Persiapan Data

Terdapat beberapa persyaratan yang digunakan untuk melakukan seleksi terhadap calon initial cabin crew pada tahap administrasi, dimana ketika memenuhi persyaratan maka akan lanjut ke tahap selanjutnya, adapun persyaratan dapat dilihat pada tabel 1.

Tabel 1 Tabel Persyaratan

| Kode Kriteria | Kriteria |
|---------------|------------------------------|
| C1 | Jenis Kelamin |
| C2 | Status Pernikahan |
| C3 | Usia |
| C4 | Pendidikan |
| C5 | Tinggi Badan |
| C6 | Buta Warna (Parsial) |
| C7 | Kemampuan Berbahasa Inggris |
| C8 | Kemampuan Berbahasa Mandarin |

Pada tabel setiap kriteria memiliki kode untuk mempermudah proses pengolahan data, masing-masing dari kriteria memiliki value dan bobot, dan setiap value memiliki bobot yang besarnya 0 sampai 3, untuk pembobotannya ditunjukkan pada tabel 2.

Tabel 2 Pembobotan Kriteria

| Kriteria | Value | Bobot |
|-----------------------------|--------------------|-------|
| Jenis Kelamin | Laki-Laki | 0 |
| | Perempuan | 1 |
| Status Pernikahan | Belum Menikah | 1 |
| | Menikah | 0 |
| | Janda/Duda | 0 |
| Usia | <18 | 0 |
| | 18-24 | 1 |
| | >24 | 0 |
| Pendidikan | SMP | 0 |
| | SMA | 1 |
| | S1 | 2 |
| | S2 | 3 |
| Tinggi Badan | <160 | 0 |
| | >=160 | 1 |
| Buta Warna (Parsial) | Ya | 0 |
| | Tidak | 1 |
| Kemampuan Berbahasa Inggris | Tidak Bisa | 0 |
| | Beginner | 0,25 |
| | Elementary | 0,50 |
| | Intermediate | 0,75 |
| | Upper Intermediate | 1 |
| | Advanced | 1,25 |
| Proficient | 1,50 | |

| | | |
|------------------------------|---------------|------|
| Kemampuan Berbahasa Mandarin | Tidak Bisa | 0 |
| | HSK Level I | 0,25 |
| | HSK Level II | 0,50 |
| | HSK Level III | 0,75 |
| | HSK Level IV | 1 |
| | HSK Level V | 1,25 |
| | HSK Level VI | 1,50 |

Pada penelitian ini memiliki dua data, data yang pertama adalah data latih dimana data ini digunakan untuk melatih metode knn, sedangkan data kedua adalah data uji dimana digunakan untuk mengetahui performa metode knn yang sudah dilatih sebelumnya ketika menemukan data baru. Pada penelitian ini terdapat 10 data latih yang akan digunakan dengan 8 atribut serta 1 label, data latih ditunjukkan pada tabel 3.

Tabel 3 Data Latih

| No | Jenis Kelamin | Status Pernikahan | Usia | Pendidikan | Tinggi Badan | Buta Warna (Parsial) | Kemampuan Berbahasa Inggris | Kemampuan Berbahasa Mandarin | Hasil |
|----|---------------|-------------------|------|------------|--------------|----------------------|-----------------------------|------------------------------|----------|
| 1 | Perempuan | Belum Menikah | 17 | SMA | 165 | Tidak | Beginner | Tidak Bisa | Berhenti |
| 2 | Perempuan | Menikah | 25 | SMA | 169 | Tidak | Elementary | HSK Level I | Berhenti |
| 3 | Perempuan | Belum Menikah | 22 | S1 | 166 | Tidak | Intermediate | HSK Level II | Lanjut |
| 4 | Perempuan | Belum Menikah | 18 | SMA | 160 | Tidak | Elementary | HSK Level II | Lanjut |
| 5 | Laki-Laki | Belum Menikah | 18 | SMA | 163 | Ya | Beginner | Tidak Bisa | Berhenti |
| 6 | Perempuan | Belum Menikah | 19 | SMA | 170 | Tidak | Intermediate | Tidak Bisa | Lanjut |
| 7 | Perempuan | Belum Menikah | 24 | SMA | 158 | Tidak | Beginner | Tidak Bisa | Berhenti |
| 8 | Perempuan | Belum Menikah | 18 | SMP | 162 | Tidak | Tidak Bisa | Tidak Bisa | Berhenti |
| 9 | Perempuan | Belum Menikah | 24 | S1 | 168 | Tidak | Intermediate | Tidak Bisa | Lanjut |
| 10 | Perempuan | Belum Menikah | 20 | SMA | 165 | Tidak | Intermediate | HSK Level III | Lanjut |

Sedangkan untuk data uji pada penelitian ini menggunakan 6 data dengan label yang sudah diketahui yang nantinya akan dibandingkan dengan hasil dari penghitungan menggunakan metode knn. Data uji pada penelitian ini ditampilkan pada tabel 4.

Tabel 4 Data Uji

| No | Jenis Kelamin | Status Pernikahan | Usia | Pendidikan | Tinggi Badan | Buta Warna (Parsial) | Kemampuan Berbahasa Inggris | Kemampuan Berbahasa Mandarin | Hasil |
|----|---------------|-------------------|------|------------|--------------|----------------------|-----------------------------|------------------------------|----------|
| 1 | Perempuan | Belum Menikah | 17 | SMA | 158 | Ya | Beginner | Tidak Bisa | Berhenti |
| 2 | Laki-Laki | Menikah | 25 | SMA | 169 | Tidak | Elementary | Tidak Bisa | Berhenti |
| 3 | Perempuan | Belum Menikah | 19 | S1 | 169 | Tidak | Intermediate | HSK Level II | Lanjut |

| | | | | | | | | |
|---|-------------------------|----|-----|-----|-------|--------------|---------------|--------|
| 4 | Perempuan Belum Menikah | 24 | S1 | 160 | Tidak | Intermediate | HSK Level III | Lanjut |
| 5 | Perempuan Belum Menikah | 20 | SMA | 167 | Tidak | Intermediate | HSK Level II | Lanjut |
| 6 | Perempuan Belum Menikah | 23 | S1 | 169 | Tidak | Advanced | HSK Level II | Lanjut |

3.2. Inisialisasi Data

Selanjutnya data yang akan melalui proses *preprocessing* ditransformasikan kedalam bentuk numerik agar dapat dilakukan penghitungan metode knn, proses inisialisasi dilakukan dengan mengubah data latih pada tabel 3 menggunakan pembobotan pada tabel , hasil inisialisasi data latih ditampilkan pada tabel 5.

Tabel 5 Inisialisasi Data Latih

| No | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|----|----|----|----|----|----|------|------|
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0,25 | 0 |
| 2 | 1 | 0 | 0 | 1 | 1 | 1 | 0,5 | 0,25 |
| 3 | 1 | 1 | 1 | 2 | 1 | 1 | 0,75 | 0,5 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 0,5 | 0,5 |
| 5 | 0 | 1 | 1 | 1 | 1 | 0 | 0,25 | 0 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 0,75 | 0 |
| 7 | 1 | 1 | 1 | 1 | 0 | 1 | 0,25 | 0 |
| 8 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| 9 | 1 | 1 | 1 | 2 | 1 | 1 | 0,75 | 0 |
| 10 | 1 | 1 | 1 | 1 | 1 | 1 | 0,75 | 0,75 |

Selain data latih, inisialisasi juga dilakukan terhadap data uji dimana mengubah data pada tabel 4 menggunakan pembobotan pada tabel 2, menghasilkan data yang ditampilkan tabel 6.

Tabel 6 Inisialisasi Data Uji

| No | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|----|----|----|----|----|----|------|------|
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0,25 | 0 |
| 2 | 0 | 0 | 0 | 1 | 1 | 1 | 0,5 | 0 |
| 3 | 1 | 1 | 1 | 2 | 1 | 1 | 0,75 | 0,5 |
| 4 | 1 | 1 | 1 | 2 | 1 | 1 | 0,75 | 0,75 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 0,75 | 0,5 |
| 6 | 1 | 1 | 1 | 2 | 1 | 1 | 1,25 | 0,5 |

3.3. Normalisasi Data

Normalisasi pada penelitian ini dilakukan pada data latih dan data uji dengan menggunakan persamaan (1), dimana terlebih dahulu mengetahui nilai terbesar dan nilai terkecil dari setiap kriteria, nilai terbesar dan nilai terkecil pada data latih ditampilkan pada tabel 7.

Tabel 7 Nilai Terbesar dan Terkecil pada Data Latih

| Kriteria | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|------------|----|----|----|----|----|----|------|------|
| <i>Max</i> | 1 | 1 | 1 | 2 | 1 | 1 | 0,75 | 0,75 |
| <i>Min</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Dengan diketahuinya nilai terbesar dan nilai terkecil pada data latih sesuai yang ditunjukkan pada tabel, selanjutnya dapat dilakukan proses normalisasi data dengan persamaan (1). Pada data latih 1 dengan kriteria 1 didapat penghitungan $(1-0)/(1-0)=1$, data latih 1 dengan kriteria 2 didapat penghitungan $(1-0)/(1-0)=1$, data latih 1 dengan kriteria 3 didapat penghitungan $(0-0)/(1-0)=0$ dan seterusnya sampai penghitungan pada data latih 10 dengan kriteria 8, hasil penghitungan noermlisasi data latih dapat dilihat pada tabel 8.

Tabel 8 Hasil Normalisasi Data Latih

| No | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|----|----|----|-----|----|----|------|------|
| 1 | 1 | 1 | 0 | 0,5 | 1 | 1 | 0,33 | 0 |
| 2 | 1 | 0 | 0 | 0,5 | 1 | 1 | 0,67 | 0,33 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0,67 |
| 4 | 1 | 1 | 1 | 0,5 | 1 | 1 | 0,67 | 0,67 |
| 5 | 0 | 1 | 1 | 0,5 | 1 | 0 | 0,33 | 0 |
| 6 | 1 | 1 | 1 | 0,5 | 1 | 1 | 1 | 0 |
| 7 | 1 | 1 | 1 | 0,5 | 0 | 1 | 0,33 | 0 |
| 8 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 10 | 1 | 1 | 1 | 0,5 | 1 | 1 | 1 | 1 |

Selain data latih yang dinormalisasi, data uji juga dilakukan proses normalisasi dengan nilai terbesar dan nilai terkecil ditampilkan pada tabel 9.

Tabel 9 Nilai Terbesar dan Terkecil pada Data Uji

| Kriteria | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|------------|----|----|----|----|----|----|------|------|
| <i>Max</i> | 1 | 1 | 1 | 2 | 1 | 1 | 1,25 | 0,75 |
| <i>Min</i> | 0 | 0 | 0 | 1 | 0 | 0 | 0,25 | 0 |

Dengan diketahuinya nilai terbesar dan nilai terkecil pada data uji sesuai yang ditunjukkan pada tabel, selanjutnya dapat dilakukan proses normalisasi data dengan persamaan (1). Pada data uji 1 dengan kriteria 1 didapat penghitungan $(1-0)/(1-0)=1$, data uji 1 dengan kriteria 2 didapat penghitungan $(1-0)/(1-0)=1$, data uji 1 dengan kriteria 3 didapat penghitungan $(0-0)/(1-0)=0$ dan seterusnya sampai penghitungan pada data uji 6 dengan kriteria 8 dengan penghitungan $(0,5-0)/(0,75-0)=0,67$, hasil penghitungan normalisasi data uji dapat dilihat pada tabel 10.

Tabel 10 Hasil Normalisasi Data Uji

| No | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|----|----|----|----|----|----|------|------|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 | 1 | 0,25 | 0 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 0,5 | 0,67 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 0,5 | 1 |
| 5 | 1 | 1 | 1 | 0 | 1 | 1 | 0,5 | 0,67 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0,67 |

3.4. Menghitung Euclidean Distance

Untuk mencari jarak antara dua titik yaitu titik pada data latih dan titik pada data uji, maka digunakan rumus Euclidean Distance dengan persamaan (2). Pada data uji 1 dengan data latih 1 didapat penghitungan $\sqrt{(1-1)^2 + (1-1)^2 + (0-0)^2 + (0,5-0)^2 + (1-0)^2 + (1-0)^2 + (0,33-0)^2 + (0-0)^2} = 1,537$. Pada data uji 2 dengan data latih 1 didapat penghitungan $\sqrt{(1-0)^2 + (1-0)^2 + (0-0)^2 + (0,5-0)^2 + (1-1)^2 + (1-1)^2 + (0,33-0,25)^2 + (0-0)^2} = 1,502$. Pada data uji 3 dengan data latih 1 didapat penghitungan $\sqrt{(1-1)^2 + (1-1)^2 + (0-1)^2 + (0,5-1)^2 + (1-1)^2 + (1-1)^2 + (0,33-0,5)^2 + (0-0,67)^2} = 1,118$. Penghitungan *euclidean distance* dilakukan sampai pada data uji 6 dengan data latih 10, hasil dari penghitungan ditampilkan pada tabel 11.

Tabel 11 Hasil Penghitungan Euclidean Distance

| Data Uji 1 | Data Uji 2 | Data Uji 3 | Data Uji 4 | Data Uji 5 | Data Uji 6 |
|------------|------------|------------|------------|------------|------------|
| 1,537 | 1,502 | 1,118 | 1,509 | 1,312 | 1,462 |
| 1,951 | 1,239 | 1,500 | 1,650 | 1,546 | 1,572 |
| 2,333 | 2,238 | 0,000 | 0,601 | 1,118 | 0,000 |
| 2,034 | 1,967 | 0,500 | 0,624 | 0,527 | 0,601 |
| 1,833 | 1,805 | 1,124 | 1,810 | 1,650 | 1,772 |
| 2,062 | 1,953 | 0,500 | 1,225 | 0,972 | 0,833 |
| 1,537 | 2,063 | 1,118 | 1,509 | 1,312 | 1,462 |
| 1,732 | 1,750 | 1,000 | 1,500 | 0,833 | 1,563 |
| 2,236 | 2,136 | 0,000 | 1,118 | 1,302 | 0,667 |
| 2,291 | 2,194 | 0,500 | 0,707 | 0,782 | 0,601 |

3.5. Menentukan Tetangga Terdekat Berdasarkan Jarak Minimum ke K

Diketahui nilai k yang ditentukan adalah 5, setelah mendapatkan nilai *euclidean distance* yang ditampilkan pada tabel 11, selanjutnya menentukan jarak tetangga terdekat dengan mengambil 5 nilai terkecil dari setiap data uji. Pada data uji ke 1 dari 10 nilai Euclidean distance nilai terkecil adalah 1,537 pada data ke 1, nilai 1,537 pada data ke 7, nilai 1,732 pada data ke 8, nilai 1,833

pada data ke 5 dan 1,951 pada data ke 2. Selanjutnya kelima data dengan nilai terkecil diambil label sesuai pada tabel 3, dan berdasarkan tabel kelima data tersebut memiliki label Berhenti. Penghitungan dilakukan sampai pada data uji ke 6, hasil penghitungan ditampilkan pada tabel 12.

Tabel 12 Tetangga Terdekat dengan K=5

| Data | Uji ke 1 | Uji ke 2 | Uji ke 3 | Uji ke 4 | Uji ke 5 | Uji ke 6 |
|--------------------|----------|----------|----------|----------|----------|----------|
| 1 | Berhenti | Berhenti | | | | |
| 2 | Berhenti | Berhenti | | | | |
| 3 | | | Lanjut | Lanjut | Lanjut | Lanjut |
| 4 | | | Lanjut | Lanjut | Lanjut | Lanjut |
| 5 | Berhenti | Berhenti | | | | |
| 6 | | Lanjut | Lanjut | Lanjut | Lanjut | Lanjut |
| 7 | Berhenti | | | | | |
| 8 | Berhenti | Berhenti | | | Berhenti | |
| 9 | | | Lanjut | Lanjut | | Lanjut |
| 10 | | | Lanjut | Lanjut | Lanjut | Lanjut |
| Jumlah Lulus | 0 | 1 | 5 | 5 | 4 | 5 |
| Jumlah Tidak Lulus | 5 | 4 | 0 | 0 | 1 | 0 |

3.6. Menetapkan Kelas

Dari hasil penghitungan pada tabel 12 dimana menentukan 5 tetangga terdekat dan dijumlahkan kelas Lanjut dan kelas Berhenti, kemudian menentukan jumlah kelas yang lebih besar dan menentukan kelas prediksinya. Hasil penetapan kelas ditampilkan pada tabel 13.

Tabel 13 Hasil Penetapan Kelas Prediksi

| Data Uji ke-n | Lanjut | Berhenti | Kelas Prediksi |
|---------------|--------|----------|----------------|
| 1 | 0 | 5 | Berhenti |
| 2 | 1 | 4 | Berhenti |
| 3 | 5 | 0 | Lanjut |
| 4 | 5 | 0 | Lanjut |
| 5 | 4 | 1 | Lanjut |
| 6 | 5 | 0 | Lanjut |

Pada tabel 13 menjelaskan bahwa data uji ke-1 dengan nilai Berhenti=5 lebih besar dari nilai Lanjut=0 maka data uji ke-1 masuk ke kelas Berhenti, data uji ke-2 nilai Berhenti lebih besar dari nilai Lanjut maka data uji ke-2 masuk ke kelas Berhenti, sedangkan untuk data uji ke-3, 4,5 dan 6 nilai Lanjut yang lebih besar dari nilai Berhenti, sehingga masuk ke kelas Lanjut.

3.7. Pengujian

Pada penelitian pengujian akurasi dilakukan dengan *Confusion Matrix*, sebelum menentukan tabel confusion matrix hasil dari kelas prediksi untuk data uji akan dibandingkan dengan kelas asli terlebih dahulu, perbandingan keduanya ditampilkan pada tabel 14.

Tabel 14 Perbandingan Kelas Asli dan Kelas Prediksi

| Data ke-n | Kelas Asli | Kelas Prediksi |
|-----------|------------|----------------|
| 1 | Berhenti | Berhenti |
| 2 | Berhenti | Berhenti |
| 3 | Lanjut | Lanjut |
| 4 | Lanjut | Lanjut |
| 5 | Lanjut | Lanjut |
| 6 | Lanjut | Lanjut |

Dari jumlah data uji sebanyak 6 data, didapat data lanjut yang terdeteksi lanjut sebanyak 4, data lanjut yang diklasifikasi berhenti sebanyak 0, data berhenti yang diklasifikasi lanjut sebanyak 0, dan data berhenti yang terdeteksi berhenti sebanyak 2. Hasil *Confusion Matrix* dapat disajikan pada tabel 15.

Tabel 15 Confusion Matrix

| Aktual | Prediksi | |
|----------|----------|----------|
| | Lanjut | Berhenti |
| Lanjut | 4 | 0 |
| Berhenti | 0 | 2 |

Pada tabel diatas *True Positive* adalah data yang berjumlah 4, *False Negative* adalah data yang berjumlah 0, *False Positive* adalah data yang berjumlah 0, dan *True Negative* adalah data yang berjumlah 2. Untuk mengetahui seberapa akurat model dalam mengklasifikasikan dengan benar, maka dihitung akurasi dengan persamaan (3), penghitungan akurasi sebagai berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{4+2}{4+2+0+0} \times 100\% \\ &= 100\% \end{aligned}$$

Dari penghitungan diatas dihasilkan nilai akurasi adalah 100%, untuk menghitung akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model menggunakan presisi yang dapat dihitung menggunakan persamaan (4), penghitungan presisi sebagai berikut:

$$\begin{aligned} \text{Presisi} &= \frac{4}{0+4} \times 100\% \\ &= 100\% \end{aligned}$$

Penghitungan diatas menghasilkan nilai presisi sebesar 100%, dan *Recall* digunakan menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi, penghitungan recall menggunakan persamaan (5), penghitungan recall dihitung sebagai berikut:

$$\begin{aligned}
 Recall &= \frac{4}{0+4} \times 100\% \\
 &= 100\%
 \end{aligned}$$

Untuk nilai *recall* yang didapat dari penghitungan diatas adalah 100%, dan hasil dari penghitungan ketiga pemodelan diatas ditampilkan pada tabel 16.

Tabel 16 Performance Vector

| Variabel | Hasil |
|----------|-------|
| Akurasi | 100% |
| Presisi | 100% |
| Recall | 100% |

4. CONCLUSION

Dari pembahasan metode k-nearest neighbor (KNN) dapat digunakan untuk mengklasifikasikan calon *initial cabin crew* pada maskapai super air jet, dengan 10 data latih yang digunakan dan 6 data uji pada penelitian ini menghasilkan 2 data dengan kelas prediksi Berhenti dan 4 data dengan kelas prediksi Lanjut, yang artinya 2 data yaitu data ke-1 dan ke-2 tidak melanjutkan ke tahap selanjutnya, sedangkan 4 data yaitu data ke-3, 4, 5 dan 6 lanjut ke tahap selanjutnya yaitu tes psikologi. Hasil dari kelas prediksi kemudian dibandingkan dengan kelas asli dan semua hasil kelasnya sama, untuk mengukur kinerja metode KNN, dilakukan pengujian menggunakan *confusion matrix* menghasilkan nilai Akurasi 100%, nilai Presisi 100% dan nilai Recall 100%.

REFERENCES

- [1] Kemenkes, "Status Wabah Corona di Indonesia Ditetapkan sebagai Bencana Nasional," kemkes.go.id, Jakarta, Mar. 15, 2020.
- [2] Puspa A. W, "Kantongi Izin AOC, Super Air Jet Segera Terbangkan 3 Pesawat A320-200," ekonomi.bisnis.com, 2021.
- [3] Murti M. W, "Super Air Jet, All-Indonesian Low-cost Carrier Launches," en.tempo.co, Jakarta, May 03, 2021.
- [4] Rahma A, "Potret Pramugari Super Air Jet, Maskapai Baru Bergaya Milenial," www.liputan6.com, May 03, 2021.
- [5] SAJ S. A. J, "Kisah Super Air Jet," superairjet.com, 2021. (accessed 2021).
- [6] Khormarudin A. N, "Teknik Data Mining: Algoritma K-Means Clustering," J. Ilmu

- Komput., pp. 1–12, 2016.
- [7] Nugraha A. R. D., Auliasari K., & Pranoto Y. A, “Implementasi Metode K-Nearest Neighbor (KNN) Untuk Seleksi Calon Karyawan Baru (Studi Kasus: BFI Finance Surabaya),” Vol 4(2), JATI (Jurnal Mhs. Tek. Inform.), 2020, pp. 14-20.
- [8] Yanosma D, Johar A, & Anggriani K, “Implementasi Metode K-Nearest Neighbor (KNN) dan Simple Additive Weighting (SAW) dalam Pengambilan Keputusan Seleksi Anggota PASKIBRAKA,” Vol 4(2), Rekursif J. Inform., 2016, pp. 222-235.
- [9] Wahidin A. J, “Implementation of K-Means Algorithm in Determining Clustering of Lion Air Pilots,” Vol 3(1), SYSTEMATICS, 2021, pp. 211-220.
- [10] Winarso D., Arribe E. E, “Seleksi Pegawai dan Dosen UMRI Berbasis E-Recruitment Menggunakan Metode K-Nearest Neighbor,” Vol 8(2), Digit. Zo. J. Teknol. Inf. dan Komun., 2017. pp. 71-80.
- [11] Danial E., Warsinah N, Metode Penulisan Karya Ilmiah. Bandung, 2009.
- [12] Narkhede S, “Understanding Confusion Matrix,” 2018.
<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
(accessed 2021).