

Comparing Naive Bayes, K-Nearest Neighbor, And Neural Network Classification Methods Of Seat Load Factor In Lombok Outbound Flights

Mega Luna Suliztia^{1*}, Achmad Fauzan^{2*}

Abstract

Classification is the process of grouping data based on observed variables to predict new data whose class is unknown. There are some classification methods, such as Naïve Bayes, K-Nearest Neighbor and Neural Network. Naïve Bayes classifies based on the probability value of the existing properties. K-Nearest Neighbor classifies based on the character of its nearest neighbor, where the number of neighbors= k , while Neural Network classifies based on human neural networks. This study will compare three classification methods for Seat Load Factor, which is the percentage of aircraft load, and also a measure in determining the profit of airline. Affecting factors are the number of passengers, ticket prices, flight routes, and flight times. Based on the analysis with 47 data, it is known that the system of Naïve Bayes method has misclassifies in 14 data, so the accuracy rate is 70%. The system of K-Nearest Neighbor method with $k=5$ has misclassifies in 5 data, so the accuracy rate is 89%, and the Neural Network system has misclassifies in 10 data with accuracy rate 78%. The method with highest accuracy rate is the best method that will be used, which in this case is K-Nearest Neighbor method with success of classification system is 42 data, including 14 low, 10 medium, and 18 high value. Based on the best method, predictions can be made using new data, for example the new data consists of Bali flight routes (2), flight times in afternoon (2), estimate of passenger numbers is 140 people, and ticket prices is Rp.700,000. By using the K-Nearest Neighbor method, Seat Load Factor prediction is high or at intervals of 80% -100%.

Keywords: Classification, K-Nearest Neighbor, Naïve Bayes, Neural Network, Seat Load Factor.

Abstrak

Klasifikasi adalah proses pengelompokkan data berdasarkan variabel yang diamati untuk memprediksi data baru yang belum diketahui kelasnya. Beberapa metode klasifikasi antara lain *Naïve Bayes*, *K-Nearest Neighbor*, dan *Neural Network*. *Naïve Bayes* bertujuan untuk klasifikasi berdasarkan nilai probabilitas dari sifat yang telah ada. *K-Nearest Neighbor* bertujuan untuk klasifikasi berdasarkan sifat tetangga terdekatnya, dimana jumlah tetangga = k , sedangkan *Neural Network* berdasarkan perilaku jaringan syaraf manusia. Penelitian ini akan membandingkan ketiga metode tersebut dalam klasifikasi *Seat Load Factor* yang merupakan persentase keterisian pesawat, dan juga ukuran dalam menentukan nilai keuntungan pada penerbangan suatu maskapai. Faktor yang mempengaruhi yaitu jumlah penumpang, harga tiket, rute, dan waktu penerbangan. Berdasarkan analisis yang dilakukan dengan 47 data diketahui bahwa sistem dengan metode *Naïve Bayes* melakukan kesalahan klasifikasi dalam 14 data sehingga tingkat akurasi sebesar 70%. Sistem dalam metode *K-Nearest Neighbor* dengan $k=5$ melakukan kesalahan klasifikasi dalam 5 data, sehingga tingkat akurasi sebesar 89%, dan sistem *Neural Network* melakukan kesalahan klasifikasi dalam 10 data sehingga tingkat

* Program Studi Statistika, Universitas Islam Indonesia,
Email: ¹16611079@students.uii.ac.id, ²achmadfauzan@uui.ac.id

akurasi sebesar 78%. Metode yang memiliki nilai akurasi tertinggi merupakan metode terbaik yang akan digunakan, yang pada kasus ini adalah metode *K-Nearest Neighbor* dengan keberhasilan sistem pada klasifikasi sebanyak 42 data, diantaranya 14 bernilai rendah, 10 sedang, dan 18 bernilai tinggi. Berdasarkan metode terbaik yang diperoleh dapat dilakukan prediksi menggunakan data baru, misalkan diketahui suatu data dengan rute penerbangan Bali (2), waktu penerbangan siang hari (2), estimasi jumlah penumpang 140 orang, dan harga tiket sebesar Rp.700.000. Dengan menggunakan metode *K-Nearest Neighbor* dihasilkan prediksi *Seat Load Factor* yang tinggi pada interval 80%-100%.

Kata kunci : Klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes*, *Neural Network*, *Seat Load Factor*

1. Pendahuluan

Salah satu alat transportasi yang menjadi andalan masyarakat karena dapat beroperasi secara efektif dari segi waktu adalah pesawat terbang. Namun dilansir dari Liputan 6 pada Januari lalu, Bulan Desember 2018 mengalami penurunan jumlah penumpang pesawat terbang secara nasional sebesar 9.75% dibanding dengan Desember 2017 [6]. Akhirnya sejumlah maskapai terpaksa mengurangi jadwal penerbangan yang ada akibat turunnya jumlah penumpang. Penurunan jumlah penumpang juga akan mengurangi *Seat Load Factor* atau SLF yang mana inilah yang akhirnya berimbas pada pendapatan maskapai yang akhirnya anjlok hingga 60%.

Machine learning merupakan cabang aplikasi dari kecerdasan buatan atau *Artificial Intelligence*. Dalam menyelesaikan masalah, *machine learning* memiliki kemampuan belajar sendiri dan menirukan perilaku manusia berupa klasifikasi dan prediksi [2]. Menurut Agnes, dkk penerapan kecerdasan buatan dalam SLF dapat meningkatkan keefisienan dan keefektifan dalam pengambilan keputusan bagi suatu maskapai penerbangan. Sehingga dari uraian tersebut peneliti akan membandingkan beberapa metode klasifikasi seperti *K-Nearest Neighbor*, *Naïve Bayes*, dan *Artificial Neural Network* terhadap nilai *Seat Load Factor* atau SLF [1].

Beberapa penelitian sebelumnya yang berkaitan dengan analisis dan studi kasus antara lain, Amin [3] menganalisis pengaruh tarif penerbangan, jumlah penerbangan, dan pendapatan perkapita dalam meningkatkan jumlah penumpang. Selanjutnya terdapat Pardede [8] yang melakukan analisis terkait faktor-faktor yang mempengaruhi permintaan penumpang maskapai penerbangan Garuda Indonesia Airlines di Provinsi Sumatera Utara. Silitonga [11] pada 2018, beliau melakukan Analisis *Load Factor* (L/F) Penumpang Pesawat Terbang Rute Palangkaraya. Windarti, dkk [13] kemudian melakukan analisis perbandingan terkait kinerja 6 algoritma klasifikasi data mining untuk prediksi masa studi mahasiswa. Selain itu Susanto [12] juga melakukan penelitian dengan membandingkan metode Algoritma *Neural Network*, *K-Nearest Neighbor*, dan *Naïve Bayes* untuk memprediksi pendonor darah potensial.

Penelitian dilakukan untuk mengetahui gambaran SLF dan faktor lain yang mempengaruhinya. Tujuan utama adalah mengetahui metode terbaik yang digunakan dalam klasifikasi SLF yang kemudian akan digunakan untuk melakukan prediksi. Dari tujuan penelitian tersebut hasil yang diperoleh dapat dimanfaatkan sebagai bahan pertimbangan dalam mengambil kebijakan untuk meningkatkan penjualan tiket guna memaksimalkan *revenue* perusahaan.

2. Metode Penelitian

Klasifikasi SLF adalah sebuah sistem yang berfungsi untuk mengelompokkan SLF berupa nilai persentase kebeberapa kelas. Dalam proses klasifikasi ini digunakan beberapa faktor lain yang dapat mempengaruhi nilai SLF antara lain waktu penerbangan (X_1), jumlah penumpang (X_2), rute penerbangan (X_3), dan harga tiket pesawat (X_4). Pada penelitian ini metode yang digunakan untuk menggambarkan data penerbangan *outbound* adalah statistika deskriptif. Dalam melakukan klasifikasi SLF, metode yang digunakan adalah *K-Nearest Neighbor*, *Naïve Bayes*, dan *Artificial Neural Network*. Hasil klasifikasi akan mengelompokkan nilai SLF kedalam kelas tertentu seperti kelas rendah, sedang, dan tinggi. Metode terbaik yang akan digunakan adalah metode yang menghasilkan tingkat akurasi tertinggi dibanding metode lainnya. Dengan menggunakan metode terbaik tersebut dapat dilakukan prediksi SLF atas suatu penerbangan. Analisis ini dilakukan menggunakan *python*.

Klasifikasi terdiri dari 2 proses yaitu *fase training*, dimana algoritma dibentuk sebagai sebuah model klasifikasi. Kemudian yang kedua adalah *fase testing* yang digunakan untuk memperkirakan tingkat akurasi dari model yang telah terbentuk. Hingga saat ini belum ada aturan baku untuk menentukan proporsi dari data *training* dan data *testing*. Namun proporsi data *training* biasanya lebih besar dibandingkan data *testing*. Para peneliti biasanya menggunakan proporsi data *training* dan data *testing* sebesar 80 : 20 [4]. Pada analisis ini data yang digunakan merupakan data sekunder yaitu data SLF penerbangan *outbound* pada maskapai Garuda *branch office* Lombok pada Januari 2019 dengan sasaran hasil analisis dapat mengrepresentasikan sifat populasi yaitu seluruh penerbangan *outbound* di Indonesia menggunakan maskapai Garuda. Dalam penelitian ini terdapat beberapa variabel yang akan digunakan, antara lain:

a. Seat Load Factor (Dependen)

SLF adalah persentase dari tempat duduk atau kapasitas penumpang pada suatu penerbangan yang berupa perbandingan antara kapasitas tempat duduk yang terjual dengan kapasitas tempat duduk yang tersedia [11].

$$SLF = \frac{\text{Jumlah Penumpang}}{\text{Kapasitas Pesawat}} \times 100 \% \quad (1)$$

SLF merupakan ukuran efisiensi untuk menentukan nilai keuntungan dari suatu penerbangan. Hal tersebut dikatakan efisien ketika pendapatan yang diperoleh berdasarkan SLF sesuai dengan peramalan yang telah dilakukan. Nilai SLF memiliki 3 kategori yaitu rendah jika SLF 31-50%, sedang jika SLF 51%-80%, dan tinggi jika SLF 81%-100%.

b. Jumlah Penumpang (Independen)

Variabel ini menggambarkan banyaknya *pax* atau penumpang yang telah pasti berada dalam penerbangan tersebut. Variabel jumlah penumpang memiliki tipe data numerik.

c. Harga Tiket (Independen)

Harga pada data SLF adalah rata-rata harga dihari penerbangan tersebut. Variabel ini memiliki tipe data numerik.

d. Rute Penerbangan (Independen)

Rute penerbangan merupakan pergerakan pesawat dari Lombok ke kota lain. Pada data ini rute yang digunakan adalah rute penerbangan dalam negeri atau domestik. Tipe data pada

variabel ini adalah kategorik dengan 5 nilai, antara lain: (1) adalah rute penerbangan dengan tujuan Jakarta, (2) adalah rute penerbangan dengan tujuan Bali, (3) adalah rute penerbangan dengan tujuan Sumbawa, (4) adalah rute penerbangan dengan tujuan Bima, dan (5) adalah rute penerbangan dengan tujuan Surabaya.

e. Waktu Penerbangan (Independen)

Waktu penerbangan memiliki tipe data kategorik. Variabel ini menggambarkan waktu yang dipilih oleh pax saat akan melakukan penerbangan. Waktu terbagi menjadi 4 nilai, yaitu: (1) pagi dengan rentang pukul 04.00 – 10.00, (2) siang dengan rentang pukul 10.00 – 14.00, (3) sore dengan rentang pukul 14.00 – 18.30, (4) malam dengan rentang pukul 18.30 – 04.00

2.1 Naïve Bayes

Metode klasifikasi yang menggunakan konsep probabilitas dengan asumsi antar kategori yang ada bersifat independen atau tidak saling bergantung pada kategori lainnya adalah metode *Naïve Bayes*. *Thomas Bayes* merupakan ilmuwan asal Inggris yang mengemukakan metode tersebut dan digunakan untuk memprediksi peluang suatu peristiwa berdasarkan peristiwa-peristiwa sebelumnya. Persamaan dari teorema Bayes tersebut adalah :

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (2)$$

Seperti persamaan diatas, nilai dari *posterior* atau peluang pada kelas X dengan sifat sampel tertentu dapat diperoleh dari peluang terjadinya kelas X sebelum terdapat sampel tersebut atau disebut *prior*, dikalikan dengan *likelihood* atau peluang terjadinya sifat sampel pada kelas X, kemudian dibagi dengan peluang terjadinya sifat sampel secara keseluruhan atau *evidence* [10].

2.2 K-Nearest Neighbor

Sebuah algoritma pelatihan untuk mengklasifikasikan suatu data baru berdasarkan kategori ketetanggaannya atau mayoritas banyaknya K data pelatihan yang terdekat dengannya (*nearest neighbor*) disebut dengan metode *K-Nearest Neighbor*. Sehingga dalam menggunakan metode ini untuk melakukan klasifikasi data baru, peneliti perlu menentukan jumlah k tetangga terdekat yang akan digunakan. Jumlah data dan ukuran dimensi data yang dimiliki akan mempengaruhi penentuan jumlah k tetangga. Semakin besar data yang digunakan maka semakin rendah jumlah k tetangga yang ditentukan. Hal tersebut dikarenakan semakin banyak k maka kemampuan untuk mengklasifikasi akan semakin rendah. Para peneliti biasanya menggunakan k 5,7 atau 9 untuk melakukan klasifikasi dengan metode *K-Nearest Neighbor*. Dekat atau jauhnya jarak titik dengan tetangganya bisa dihitung dengan menggunakan jarak *euclidean* seperti persamaan dibawah ini :

$$D(x, y) = \sqrt{\sum_{k=1}^{k_n} (a_k - b_k)^2} \quad (3)$$

a = titik yang telah diketahui kelasnya, b = titik baru, dan D = Jarak. D(x,y) adalah jarak antara titik yang telah diketahui klasifikasinya (a) dan titik baru (b). Jarak antara titik baru dengan titik data training akan dihitung dan ditentukan jumlah tetangga (k) dalam analisis. Kemudian titik baru tersebut diprediksi kedalam klasifikasi sesuai dengan kelas terbanyak dari titik-titik [7].

2.3 Artificial Neural Network

Artificial Neural Network (ANN) atau jaringan syaraf tiruan merupakan salah satu metode yang bekerja dengan cara meniru perilaku jaringan syaraf (*neuron*) atau cara kerja otak manusia.

Metode ini dirancang untuk memecahkan masalah yang tidak dapat diselesaikan dengan komputasi konvensional. Jaringan-jaringan tersebut terdiri atas sekumpulan *neuron* yang saling berinteraksi untuk memecahkan kasus seperti klasifikasi dan prediksi. Dalam menggunakan metode ini, suatu *neuron* dapat memiliki beberapa *input* untuk diolah dan hanya memiliki satu *output* yang merupakan hasil akhir. Struktur *artificial neural network* terdiri atas 3 lapisan yaitu lapisan *input*, lapisan *hidden* atau lapisan antara *input* dan *output*, dan lapisan terakhir adalah lapisan [5].

2.4 Pengukuran Akurasi Klasifikasi

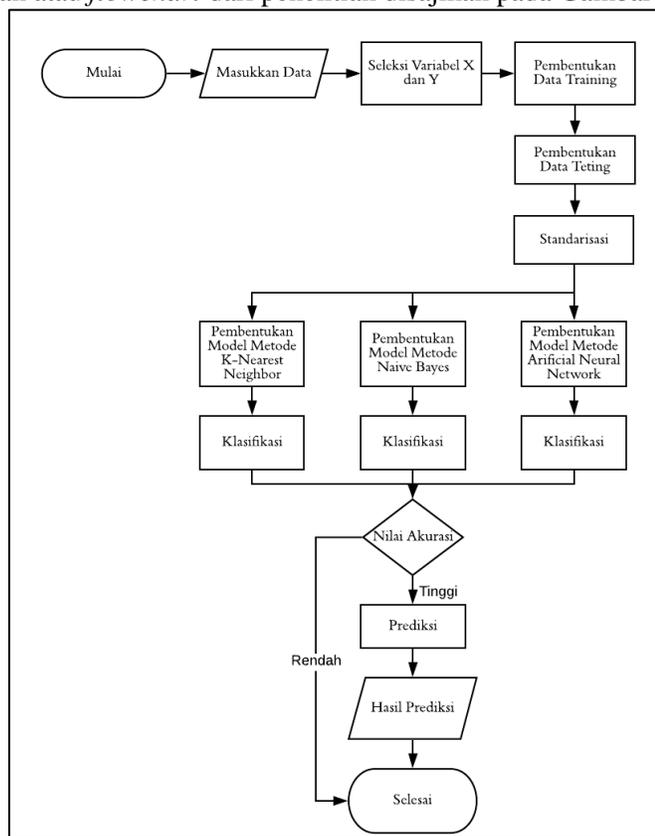
Akurasi merupakan suatu ukuran untuk mengetahui seberapa baik hasil pengukuran terhadap angka aktual yang ada. Nilai akurasi yang berupa persentase kebenaran dapat diperoleh dari perbandingan antara jumlah pengukuran yang benar terhadap data aktual dengan jumlah seluruh data yang digunakan. Rumus akurasi adalah :

$$\text{Akurasi} = \frac{\text{Jumlah nilai prediksi benar}}{\text{Jumlah data}} \times 100\% \quad (4)$$

Saat ini tidak ada batasan atau *range* untuk nilai akurasi dalam menjelaskan kebaikan model. Namun, nilai akurasi akan dikatakan baik ketika mendekati nilai 100% [9].

2.5 Tahapan Analisis

Dalam melakukan penelitian yang baik, diperlukan suatu tahapan agar penelitian terarah dan sesuai struktur. Tahapan atau *flowchart* dari penelitian disajikan pada Gambar 1.



Gambar 1. FlowChart Analisis

Berdasarkan *flowchart* atau diagram alir pada gambar 1, diketahui langkah-langkah yang harus dilalui dalam melakukan analisis berikut ini.

1. Peneliti memasukkan data pada *python* dengan format *.csv*.
2. Kemudian peneliti memilih variabel yang akan menjadi variabel independen dengan cara mengeluarkan variabel dependen dari data. Lalu peneliti membuat data baru yang berisikan data yang telah dikeluarkan sebagai variabel dependen.
3. Selanjutnya peneliti memecah data menjadi dua yaitu data *training* dan data *testing*. Data *training* akan digunakan untuk membentuk model klasifikasi pemodelan
4. Kemudian peneliti akan membentuk data *testing* yang akan digunakan untuk menguji kebaikan dari model yang telah dibentuk.
5. Selanjutnya peneliti melakukan standarisasi untuk menyamakan skala atau satuan dari setiap variabel kedalam bentuk standar.
6. Setelah itu peneliti akan melakukan analisis klasifikasi menggunakan metode *Naïve Bayes*, *K- Nearest Neighbor* dan *Artificial Neural Network*. Untuk melakukan klasifikasi dengan ketiga metode ini tidak ada asumsi yang harus dipenuhi. Kemudian dilakukan pembentukan model dari data *train*. Untuk metode *K- Nearest Neighbor*, peneliti harus menentukan jumlah ketetanggaan yang akan digunakan yang pada kasus ini digunakan pendekatan dengan $k = 5$.
7. Selanjutnya lakukan prediksi atas daya *Y test*. Kemudian data tersebut akan dibandingkan dengan data *Y test* aktual untuk diketahui klasifikasinya.
8. Kemudian untuk menentukan metode terbaik, peneliti akan membandingkan nilai akurasi dari ketiga metode tersebut, metode terbaik adalah metode yang memiliki nilai akurasi tertinggi. Untuk nilai akurasi yang lebih rendah, proses klasifikasi akan terhenti. Sedangkan untuk nilai akurasi yang lebih tinggi akan digunakan untuk prediksi.
9. Saat melakukan prediksi, peneliti akan memasukkan data baru sesuai dengan variabel independen yang ada. Kemudian diprediksi menggunakan metode terbaik sehingga akan diketahui nilai klasifikasi yang sesuai dengan ciri-cirinya.
10. Kemudian tahapan analisis akan selesai.

3. Hasil dan Pembahasan

3.1 Statistika Deskriptif

Berdasarkan analisis, diperoleh gambaran secara umum pada Gambar 2 sampai Gambar 4.



Gambar 2. Pie Chart Seat Load Factor



Gambar 3. Pie Chart Rute Penerbangan



Gambar 4. Pie Chart Waktu Penerbangan

Gambar 2 memberikan informasi bahwa persentase *Seat Load Factor* memiliki nilai yang berbeda-beda, dimana 43.8% dari seluruh penerbangan *outbound* menghasilkan *Seat Load Factor* dengan kategori tinggi, 22.7% termasuk dalam kategori sedang, dan sisanya adalah *Seat Load Factor* dengan kategori rendah. Dari *pie chart* diatas maka dapat disimpulkan bahwa sejauh ini perusahaan mampu menghasilkan *Seat Load Factor* yang tinggi. Gambar 3 memberikan informasi bahwa proporsi rute penerbangan memiliki persentase yang berbeda-beda, dimana 26.6% dari seluruh penerbangan *outbound* memiliki tujuan ke Jakarta, 27.9% penerbangan dengan tujuan Bali, 21% adalah penerbangan dengan tujuan Sumbawa, 11.6% dengan tujuan Bima, dan 12.9% dengan tujuan Surabaya. Dari *pie chart* diatas maka dapat disimpulkan bahwa pesawat paling sering beroperasi dengan tujuan Bali, namun tidak dengan Bima.

Sementara Gambar 4 memberikan informasi bahwa proporsi waktu penerbangan memiliki persentase yang berbeda-beda, dimana 42.5% dari seluruh penerbangan *outbound* berada di pagi hari, 31.3% berada di siang hari, 21.5% adalah penerbangan sore hari, dan 4.7% penerbangan malam hari. Dari *pie chart* diatas maka dapat disimpulkan bahwa penerbangan pada malam hari tergolong sedikit dan didominasi oleh penerbangan pagi hari.

Disamping deskripsi dari SLF, rute penerbangan dan waktu penerbangan, pada Gambar 5 dan Gambar 6 diberikan terkait rata-rata jumlah penumpang dan rata-rata harga tiket pesawat.



Gambar 5. Bar Chart Rata- Rata Jumlah Penumpang



Gambar 6. Bar Chart Rata- Rata Harga Tiket Pesawat

Gambar 5 memberikan informasi bahwa rata-rata penumpang yang berbeda-beda, dimana penerbangan tujuan Jakarta memiliki rata-rata mendekati 120 penumpang, kemudian menurun untuk tujuan Bali yang berkisar pada 90 penumpang, serta Sumbawa dan Bima yang masing-masing memiliki rata-rata penumpang berkisar antara 50 dan 30. Sedangkan untuk tujuan Surabaya, rata-rata penumpang pesawat adalah 70. Sehingga dapat disimpulkan bahwa adanya perbedaan permintaan penumpang yang dipengaruhi oleh rute penerbangan.

Gambar 6 memberikan informasi bahwa rata-rata harga tiket pesawat berdasarkan rute penerbangan memiliki nilai yang berbeda-beda, dimana penerbangan tujuan Jakarta memiliki harga tiket termahal dengan rata-rata Rp.1.600.000 , kemudian Bali dan Sumbawa memiliki harga rata-rata yang hampir sama yaitu diatas Rp.400.000. Bima memiliki harga rata-rata yang lumayan tinggi yaitu lebih dari Rp.800.000, sedangkan Surabaya lebih dari Rp. 1.000.000.

Selanjutnya adalah analisis klasifikasi, dimana sebuah model klasifikasi dibentuk berdasarkan data yang sudah ada, kemudian model digunakan untuk melakukan klasifikasi dan dilanjutkan untuk prediksi data baru yang belum ada. Dalam kasus ini peneliti akan mengklasifikasi nilai *Seat Load Factor* dari penerbangan *outbond* yang dikategorikan menjadi 3 nilai yaitu rendah, sedang,

dan tinggi dengan variabel dependen yaitu pengaruh waktu penerbangan, jumlah penumpang, rute penerbangan dan harga tiket. Banyaknya data (N) sebesar 233.

Pada umumnya, data dalam klasifikasi akan dibagi menjadi *training set* dan *testing set*. Data yang ada pada *testing set* biasanya lebih besar dari yang ada dalam *training set*. Data *training* digunakan untuk membentuk model klasifikasi. Model tersebut merupakan representasi yang digunakan untuk melakukan prediksi kelas pada data baru. Sedangkan data *testing* digunakan untuk menguji kebaikan dari model yang telah dibentuk dengan benar. Maka peneliti membagi data menjadi *training set* dan *testing set* dengan proporsi 0.8 : 0.2, 186 data tergolong dalam *training set* dan 47 lainnya adalah *testing set*. Pembagian dilakukan secara *random*, sehingga diperoleh hasil berdasarkan kategori *Seat Load Factor* berikut ini:

Tabel 1. Proporsi Data *Train* dan *Test*

Kategori	Data <i>Train</i>	Data <i>Test</i>	Total
1	63	15	78
2	40	13	53
3	83	19	102
Total	186	47	233

Selanjutnya melakukan standarisasi data untuk menyamakan skala pada setiap variabel ke dalam bentuk standar, sehingga dapat dianalisis dengan seragam. Standarisasi data dilakukan terhadap variabel independen. Standarisasi tidak dilakukan pada variabel dependen dikarenakan rentang yang ada pada variabel tersebut masih cukup kecil.

3.2 Hasil Klasifikasi

Metode *K-Nearest Neighbor* bergantung pada nilai ketetanggaan atau k yang sudah ditentukan dalam melakukan klasifikasi. Untuk menentukan nilai k terbaik digunakan nilai akurasi tertinggi, opsi nilai k yang akan digunakan adalah 5,7,9. Maka di dapatkan hasil sebagai berikut.

Tabel 2. Nilai Ketetanggaan

K	Accuracy
5	0.89
7	0.87
9	0.83

Setelah melakukan percobaan di dapatkan hasil untuk nilai k terbaik adalah 5 karena memiliki nilai akurasi 0.89 atau 89% dan merupakan nilai akurasi tertinggi dibandingkan k yang lain.

Metode klasifikasi menggunakan *Naïve Bayes* merupakan metode yang menggunakan probabilitas dalam proses klasifikasinya. Sedangkan Metode klasifikasi *Artificial Neural Network* merupakan metode yang menggunakan sistem syaraf tiruan dalam pembentukan model. Setelah melakukan pembentukan model, dan mencari nilai prediksi untuk nilai y *test*, maka diperoleh klasifikasi yang diukur dalam *confusion matrix*. *Confusion matrix* merupakan ukuran yang digunakan untuk menganalisis seberapa baik model klasifikasi dapat mengenali sampel aktual dari kategori / kelas yang berbeda.

Tabel 3. Klasifikasi dengan metode *Naïve Bayes*, *K-Nearest Neighbor*, dan *Neural Network*

<i>Actual</i>	<i>Naïve Bayes</i>			<i>K-Nearest Neighbor</i>			<i>Neural Network</i>		
	<i>Predicted Class</i>								
	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
Rendah	13	2	0	14	1	0	14	1	0
Sedang	7	3	3	2	10	1	2	5	6
Tinggi	0	2	17	0	1	18	0	1	18

Berdasarkan **Tabel 3** diatas dapat di interpretasikan menggunakan metode *Naïve Bayes*, model memprediksi perolehan *Seat Load Factor* rendah sebanyak 20 kali, namun hanya 13 yang terklasifikasi dengan benar atau tepat pada kategori rendah dan 7 lainnya berkategori sedang. Sedangkan menggunakan metode *K-Nearest Neighbor* dan *Neural Network*, model memprediksi perolehan *Seat Load Factor* rendah sebanyak 16 kali, 14 diantaranya terklasifikasi dengan benar atau tepat pada kategori rendah dan 2 lainnya berkategori sedang. Metode *Naïve Bayes*, model memprediksi perolehan *Seat Load Factor* sedang sebanyak 7 kali, namun hanya 3 kali yang terklasifikasi dengan benar atau tepat pada kategori sedang dan masing-masing 2 lainnya berada di kategori rendah dan tinggi. Dengan metode *K-Nearest Neighbor*, model memprediksi perolehan *Seat Load Factor* sedang sebanyak 12 kali, 10 kali diantaranya terklasifikasi dengan benar atau tepat pada kategori sedang dan masing-masing 1 lainnya berada di kategori rendah dan tinggi. Sedangkan dengan metode *Neural Network*, model memprediksi perolehan *Seat Load Factor* sedang sebanyak 7 kali, 5 kali diantaranya terklasifikasi dengan benar atau tepat pada kategori sedang dan masing-masing 1 lainnya berada di kategori rendah dan tinggi.

Sementara Menggunakan metode *Naïve Bayes*, model memprediksi perolehan *Seat Load Factor* tinggi sebanyak 20 kali, namun hanya 17 kali yang terklasifikasi dengan benar atau tepat pada kategori tinggi dan 3 kali lainnya berada di kategori sedang. Dengan metode *K-Nearest Neighbor*, model memprediksi perolehan *Seat Load Factor* tinggi sebanyak 19 kali, 18 kali diantaranya terklasifikasi dengan benar atau tepat pada kategori tinggi dan 1 lainnya berada di kategori sedang. Sedangkan dengan metode *Neural Network*, model memprediksi perolehan *Seat Load Factor* tinggi sebanyak 24 kali, 18 kali diantaranya terklasifikasi dengan benar atau tepat pada kategori tinggi dan 6 lainnya berada di kategori sedang

Berdasarkan pernyataan klasifikasi diatas, klasifikasi menggunakan metode *Naïve Bayes* memiliki tingkat kesalahan yang lebih tinggi dibandingkan dengan metode *K-Nearest Neighbor* dan *Neural Network*. Hal tersebut dapat dilihat dari banyaknya kesalahan klasifikasi saat diprediksi oleh sistem yang dibandingkan dengan data aktual yang ada.

3.3 Pemilihan Metode Terbaik

Pemilihan metode terbaik untuk klasifikasi dapat dilakukan berdasarkan nilai akurasi yang terbentuk. Semakin besar nilai akurasi maka semakin baik metode tersebut untuk digunakan dalam klasifikasi.

- **Nilai akurasi metode *Naïve Bayes***

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah nilai prediksi benar}}{\text{Jumlah data}} \times 100\% \\ &= \frac{13 + 3 + 17}{47} \times 100\% \\ &= 70\% \end{aligned}$$

- **Nilai akurasi metode *K-Nearest Neighbor***

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah nilai prediksi benar}}{\text{Jumlah data}} \times 100\% \\ &= \frac{14 + 10 + 18}{47} \times 100\% \\ &= 89\% \end{aligned}$$

- **Nilai akurasi metode *Neural Network***

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah nilai prediksi benar}}{\text{Jumlah data}} \times 100\% \\ &= \frac{14 + 5 + 18}{47} \times 100\% \\ &= 79\% \end{aligned}$$

Pada kasus diatas, diketahui bahwa nilai akurasi dengan metode *Naïve Bayes* adalah 70%, artinya model dengan metode *Naïve Bayes* mampu memprediksi klasifikasi data dengan tingkat kebenaran 70%. sedangkan dengan metode *K-Nearest Neighbor* diperoleh nilai akurasi sebesar 89%, artinya model dengan metode *K-Nearest Neighbor* mampu memprediksi klasifikasi data dengan tingkat kebenaran 89%. Metode *Neural Network* menghasilkan tingkat akurasi sebesar 79%, artinya model dengan *Neural Network* mampu memprediksi klasifikasi data dengan tingkat kebenaran 79%. Sehingga dapat disimpulkan bahwa metode algoritma *K-Nearest Neighbor* lebih baik dibandingkan metode *Naïve Bayes* dan *Neural Network* untuk mengklasifikasikan data *Seat Load Factor* pada penerbangan *outbound*.

3.4 Prediksi

Setelah diketahui metode terbaik yang dapat digunakan untuk klasifikasi, maka metode tersebut juga dapat digunakan untuk memprediksi klasifikasi pada data baru. Contoh studi kasus: Akan diadakan penambahan frekuensi penerbangan dengan tujuan Lombok – Bali atas pertimbangan banyaknya permintaan penumpang. Faktor- faktor yang harus diperhatikan untuk mengetahui klasifikasi *Seat Load Factor* telah dipertimbangkan sebagai berikut:

- Rute untuk penambahan penerbangan adalah Bali (Rute =2)
- Waktu penerbangan baru yang akan dibentuk adalah siang hari (Waktu = 2)
- Rata-rata jumlah penumpang dengan tujuan bali sebesar 140 orang
- Harga tiket yang akan dijual sebesar Rp. 700.000

Data diatas, kemudian akan dimasukkan kedalam program untuk dilakukan prediksi. Setelah dianalisis menggunakan metode *K-Nearest Neighbor*, maka diperoleh prediksi *Seat Load Factor*

untuk penerbangan baru dengan beberapa pertimbangan diatas akan bernilai tinggi, yang artinya penjualan Seat Load Factor akan berada di interval 81% - 100%.

4. Kesimpulan

Berdasarkan hasil yang telah dijelaskan oleh penulis pada bab pembahasan sebelumnya, maka diperoleh beberapa kesimpulan sebagai berikut :

1. Diketahui 43.8% nilai SLF dari seluruh penerbangan bernilai tinggi dengan rute penerbangan terbanyak menuju Bali sebesar 27.9% dan 42.5% dari seluruh penerbangan terjadi pada pagi hari dan juga rata-rata penumpang terbanyak menuju Jakarta sebanyak lebih dari 100 dengan rata-rata harga tiket mencapai Rp.1.600.000.
2. Berdasarkan hasil penelitian, diperoleh hasil dari metode klasifikasi *Naïve Bayes* nilai prediksi data yang sesuai dengan data aktual berjumlah 33 kali dimana 13 diantaranya bernilai rendah, 3 sedang, dan 17 bernilai tinggi. s metode klasifikasi *K-Nearest Neighbor* dan nilai $k=5$, diketahui bahwa prediksi data yang sesuai dengan data aktual berjumlah 42 dimana 14 diantaranya bernilai rendah, 10 sedang, dan 18 bernilai tinggi, metode klasifikasi *Neural Network*, diketahui bahwa prediksi data yang sesuai dengan data aktual berjumlah 37 dimana 14 diantaranya bernilai rendah, 5 sedang, dan 18 bernilai tinggi sementara nilai akurasi yang diperoleh untuk masing-masing metode dengan metode *Naïve Bayes* sebesar 70%, *K-Nearest Neighbor* sebesar 89% dan *Neural Network* sebesar 79%, maka dapat disimpulkan bahwa metode *K – Nearest Neighbor* adalah metode terbaik untuk digunakan dalam memprediksi klasifikasi suatu data karena memiliki tingkat akurasi tertinggi dibandingkan 2 metode lainnya.

5. Saran

Penelitian selanjutnya dapat melakukan perbandingan dengan metode – metode klasifikasi lainnya dan menggunakan beberapa perbandingan proporsi data *training* dan *testing* yang berbeda ataupun melakukan analisis lain dengan data yang sama seperti analisis regresi ordinal untuk mengetahui faktor-faktor yang mempengaruhi nilai SLF.

Daftar Pustaka

- [1] Agnes, K. (2010). *Analisis Dan Perancangan Business Intelligence Untuk Seat Load Factor Pada Pt. Garuda Indonesia*. DKI Jakarta: Universitas Bina Nusantara.
- [2] Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia* .
- [3] Amin, A. P. (2013). Analisis Pengaruh Tarif Penerbangan, Jumlah Penerbangan dan Pendapatan Perkapita dalam Meningkatkan Jumlah Penumpang. *Jurnal Bisnis Strategi*, 56-60.
- [4] Badu, Z. S. (2016). Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Dana Desa. *Jurnal Informatika, Gorontalo*, 8.

- [5] Effendi, A. (2013). *Penggunaan Arificial Neural Network Untuk Mendeteksi Kelainan Mata Miopi pada Manusia dengan Metode Backpropagation*. Malang: Universitas Negeri Maulana Malik Ibrahim.
- [6] Liputan6. (2019, Januarii 15). *Liputan 6*. Dipetik Maret 16, 2019, dari Harga Tiket Mahal, Jumlah Penumpang Pesawat Turun di 2018: <https://www.liputan6.com/bisnis/read/3871737/harga-tiket-mahal-jumlah-penumpang-pesawat-turun-di-2018>
- [7] Naufal, M. A. (2017). *Implementasi Metode Klasifikasi K-Nearest Neighbor untuk Pengenalan Pola Batik Lampung*. Bandar Lampung: Universitas Lampung.
- [8] Perdede, R. P. (2013). *Analisis Faktor-Faktor yang Mempengaruhi Permintaan Penumpang Maskapai Penerbangan Garuda Indonesia Airlines di Provinsi Sumatera Utara*. Medan: Universitas Negeri Medan .
- [9] Rahman, M. A. (2018). Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naive Bayes Untuk Klasifikasi Kualitas Air Bersih. *Pengembangan Teknologi Informasi dan Ilmu Komputer* , 6348.
- [10] Saleh, A. (2015). *Penerapan Data Mining dengan Metode Klasifikasi Naive Bayes untuk Memprediksi Kelulusan Mahasiswa dalam Mengikuti English Proficiency Test*. Medan : Universitas Potensi Utama.
- [11] Silitonga, S. P. (2018). Analisis Load Factor (L/F) Penumpang Pesawat Terbang Rute Palangkaraya- Jakarta. *Jurnal Teknika* .
- [12] Susanto, W. E. (2016). Komparasi Algoritma Neural Network, K-Nearest Neighbor dan Naive Bayes untuk Memprediksi Pendonor Darah Potensial. *Journal Speed-Sentra Penelitian Engineering dan Edukasi* , 20.
- [13] Windarti, M. A. (2019). Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa. *Jurnal Telematika Vol. 1, No. 1 Februari 2019*, 26.