

KOMPARASI ALGORITMA DECISION TREE, RANDOM FOREST, SVM DAN K-NN DALAM KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN

Muzakki Hafizh Setiono

Informatika
Universitas Nusa Mandiri
<https://nusamandiri.ac.id>
zaeroblitz@gmail.com



Abstract— In this era of globalization has changed various paradigms in the business sector which has an impact on increasing competition between airlines. Along with the growing interest of potential customers, every airline has taken various ways to improve the type or standard of service quality. In evaluating the quality of services provided by airlines from a business perspective, this aspect of service must be evaluated based on indicators of customer satisfaction. Data mining technique is one method that can be used in determining the level of customer satisfaction. This evaluation can be done by using a classification model. The classification algorithms used in this study include: Decision Tree, Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). The stages in this research are data analysis, data preprocessing, model making, model testing and evaluating the results of the model made. From the results of the research conducted, it can be seen that Random Forest is the best classification algorithm in this case study, with an accuracy rate of 96%. As for Decision Tree, Support Vector Machine (SVM) and K-Nearest Neighbor each have an accuracy rate of 94%, 95% and 93%.

Keywords: Airline Passenger Satisfaction, Data Mining, Decision Tree, Random Forest, K-NN, SVM

Abstrak— Pada era globalisasi ini telah mengubah berbagai paradigma dalam sektor bisnis yang berdampak pada meningkatnya persaingan antar maskapai penerbangan. Seiring dengan minat calon konsumen yang terus tumbuh, berbagai cara telah dilakukan oleh setiap maskapai penerbangan dalam meningkatkan tipe atau standar kualitas pelayanan. Dalam mengevaluasi kualitas layanan yang diberikan oleh maskapai penerbangan dari perspektif bisnis, aspek layanan ini harus dievaluasi berdasarkan indikator kepuasan konsumen. Teknik *data mining* adalah salah satu metode yang dapat digunakan dalam menentukan tingkat kepuasan konsumen. Evaluasi ini dapat dilakukan dengan menggunakan model klasifikasi. Algoritma klasifikasi yang digunakan dalam penelitian ini antara lain: *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)*. Adapun tahapan dalam penelitian ini melakukan analisis data, data preprocessing, pembuatan model, melakukan uji model dan mengevaluasi hasil dari model yang dibuat. Dari hasil penelitian yang dilakukan dapat diketahui bahwa *Random Forest* merupakan algoritma klasifikasi yang paling baik dalam studi kasus ini, dengan nilai tingkat akurasi yaitu 96%. Adapun untuk *Decision Tree*, *Support Vector Machine (SVM)* dan *K-Nearest Neighbor* masing-masing memiliki tingkat akurasi 94%, 95% dan 93%.

Kata kunci: Kepuasan Maskapai Penerbangan, Data Mining, Decision Tree, Random Forest, K-NN, SVM

PENDAHULUAN

Pada zaman yang sudah modern ini seiring dengan meningkatnya jumlah penduduk dan mobilitas manusia, maka dibutuhkan suatu alat atau sarana penunjang yang dapat memenuhi kebutuhan manusia dalam menjalankan aktivitas sosialnya. Sarana penunjang tersebut antara lain adalah layanan transportasi. Transportasi merupakan

unsur terpenting dalam pembangunan suatu daerah maupun negara yang dapat mempermudah arus mobilitas manusia, barang, maupun informasi yang dapat membantu proses pembangunan ekonomi, pertumbuhan sektor industri dan perkembangan hidup masyarakat (Siti Fatimah, 2019).

Dalam perkembangannya, transportasi telah menjadi kebutuhan ekonomi sekunder masyarakat. Transportasi juga telah membawa dampak pada

pembangunan wilayah secara menyeluruh terutama dalam menghubungkan suatu wilayah dengan berbagai wilayah lainnya (aksebilitas). Di mana dalam perkembangannya transportasi telah mengubah hampir seluruh wajah dunia menjadi jembatan pembangunan suatu wilayah, hingga pada suatu daerah maupun negara tersebut transportasi menjadi suatu kebutuhan yang tidak dapat diabaikan lagi (Rudi & Asrul, 2018).

Untuk menunjang pergerakan antar wilayah yang terdiri atas berbagai pulau-pulau besar dan kecil di dunia, terdapat 3 jenis transportasi dalam mendukung aktivitas dan mobilitas masyarakat. Ketiga jenis transportasi tersebut antara lain adalah transportasi darat, laut dan udara. Perjalanan udara merupakan salah satu moda transportasi yang mengalami perkembangan signifikan di dunia saat ini. Penggunaan perjalanan udara dipilih karena dibandingkan dengan moda transportasi lainnya memungkinkan aktivitas dan mobilitas manusia dapat dilakukan lebih cepat dan lebih mudah, tanpa memerlukan tambahan waktu yang signifikan (Widjaja, 2017).

Pada era globalisasi ini telah mengubah berbagai paradigma dalam sektor bisnis yang berdampak pada meningkatnya persaingan antar maskapai penerbangan. Seiring dengan minat calon konsumen yang terus tumbuh, berbagai cara telah dilakukan oleh setiap maskapai penerbangan dalam meningkatkan tipe atau standar kualitas pelayanan agar dapat bertahan dari beberapa pesaing yang semakin kompetitif. Setiap maskapai penerbangan bersaing dalam menghadirkan berbagai kemudahan bagi konsumen dalam menggunakan jasa yang diberikan. Semakin baik metode dan standar kualitas layanan maskapai, semakin besar kepuasan dan minat konsumen terhadap penggunaan layanan oleh calon konsumen (Gultom, Marlita, & Wardana, 2018)(Lestari, 2019).

Dalam mengevaluasi kualitas layanan yang diberikan oleh maskapai penerbangan dari perspektif bisnis, aspek layanan ini harus dievaluasi berdasarkan indikator kepuasan konsumen. Teknik *data mining* adalah salah satu metode yang dapat digunakan dalam menentukan tingkat kepuasan konsumen. Evaluasi ini dapat dilakukan dengan menggunakan model klasifikasi. Pemanfaatan model klasifikasi merupakan salah satu strategi data mining yang dapat dimanfaatkan untuk memprediksi tingkat kepuasan konsumen. Data yang diperoleh dapat diklasifikasikan dengan menggunakan berbagai strategi atau algoritma pemodelan klasifikasi (Yoga Religia & Amali, 2021). Decision Tree merupakan algoritma Tree biasa dipakai untuk pengenalan pola statistik. Decision Tree terbuat dari tiga simpul yaitu leaf, lalu terdiri juga dari simpul root yang merupakan titik awal dari suatu decision tree, dan yang terakhir adalah

simpul perantara yang berhubungan dengan suatu pengujian (Puspita & Widodo, 2021)

Random Forest merupakan salah satu metode berbasis klasifikasi dan regresi dimana terdapat proses agregasi pohon keputusan. Metode ini digunakan karena menghasilkan kesalahan yang lebih rendah, memberikan akurasi yang bagus dalam klasifikasi dan dapat menangani data pelatihan yang jumlahnya sangat besar. (Primajaya & Sari, 2018).

Metode naïve bayes merupakan algoritma yang mampu menerima input dalam bentuk apapun dan memiliki kecepatan dalam memproses suatu data (Yuliana, Paradise, & Kusri, 2021). Sedangkan KNN merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Cahyanti, Rahmayani, & Husniar, 2020)

Penelitian sebelumnya dalam pembuatan model klasifikasi menggunakan salah satu algoritma yaitu, *Naïve Bayes*. Dan hasil akurasi yang didapatkan dari penelitian sebelumnya sebesar 81,466% dengan data training yang digunakan sebesar 90% dan data testing yang digunakan sebesar 10% (Wijayanto, Bernardo, & Pamungkas, 2021). Menurut penulis, hasil tingkat akurasi yang didapatkan dari penelitian sebelumnya masih dinilai belum cukup memadai, maka dari itu penulis melakukan penilitan dengan menggunakan algoritma klasifikasi yang berbeda yaitu *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)* dalam studi kasus klasifikasi kepuasan penumpang maskapai penerbangan.

BAHAN DAN METODE

Penelitian ini terdiri dari beberapa tahap yang terdiri dari pengumpulan data, analisis data, *data preprocessing*, pembuatan model, pengujian model dan evaluasi.

Pengumpulan Data

Adapun sumber data yang digunakan dalam penilitan ini berasal dari *Kaggle* dengan judul *Airline Passenger Satisfaction*, dengan total data sebanyak 129.880 baris data yang sudah terbagi menjadi data *training* dan data *testing*.

Analisis Data

Tahap analisis data ini dilakukan untuk menganalisis *dataset* yang akan digunakan. Dengan analisis data ini dapat mengetahui apakah data memiliki *noise* atau tidak. Jika terdapat *noise* maka data harus dilakukan *preprocessing* agar data yang digunakan merupakan data yang valid dan berguna.

Data Preprocessing

Setelah dilakukan tahap analisis data, maka hal yang harus dilakukan adalah melakukan *preprocessing* atau membersihkan data dari berbagai macam *noise* yang telah ditemukan dari tahap analisis dari *dataset* yang digunakan. Hal ini bertujuan agar *dataset* yang akan digunakan memiliki *dataset* yang baik, beragam dan layak digunakan. Dimana hal ini merupakan suatu faktor yang dapat menentukan hasil dari *data mining*.

Pembuatan Model

Setelah data diolah dan ditransformasikan, selanjutnya akan dibuat sebuah model dengan melakukan pengaturan pemodelan untuk mengoptimalkan hasil. Dan di dalam setiap model yang akan dibangun akan dilakukan pengecekan beberapa *hyperparameter* dengan menggunakan *Grid Search CV*. Hal ini dilakukan agar dapat mengetahui parameter yang tepat pada setiap model algoritma dan dapat memberikan hasil tingkat akurasi yang lebih akurat.

Pengujian Model

Hasil dari setiap model algoritma akan dievaluasi kemampuannya untuk melakukan prediksi terhadap *data testing* yang telah disiapkan, untuk menentukan algoritma mana yang lebih unggul. Sehingga menghasilkan algoritma yang terbaik dan efektif untuk diterapkan pada *dataset* tersebut dalam melakukan klasifikasi.

Evaluasi

Pada langkah ini, pola atau informasi yang ditemukan dalam proses data mining akan dievaluasi, untuk mengetahui kinerja dari algoritma yang telah dilakukan mengetahui kualitas dan efektifitas sebuah model, menentukan apakah model dalam kenyataannya telah mencapai tujuan.

HASIL DAN PEMBAHASAN

Atribut & Label

Data *Airline Passenger Satisfaction* merupakan data yang berisi survei kepuasan penumpang maskapai penerbangan di dunia. Data ini mempunyai 22 atribut, 1 buah label data dan terdiri dari 129.880 record. Atribut yang terdapat pada data yang digunakan ada yang bersifat kategorik dan numerik. Label data yang digunakan pada dataset terdapat pada atribut "*Satisfaction*". Data selanjutnya akan diolah melalui bahasa pemrograman Python dan dengan menggunakan Jupyter Notebook sebagai IDE (Integrated Development Environment) nya.

Setiap atribut dan label yang digunakan pada penelitian yang akan dilakukan ini dapat dilihat pada Tabel dibawah ini.

Tabel 1. Atribut & Label Dataset

	Keterangan	Atribut/ Label
<i>Gender</i>	Jenis kelamin penumpang (Wanita/Pria)	Atribut
<i>Customer Type</i>	Jenis pelanggan (Pelanggan setia, pelanggan tidak setia)	Atribut
<i>Age</i>	Usia penumpang	Atribut
<i>Type of Travel</i>	Tujuan penerbangan penumpang (Perjalanan Pribadi, Perjalanan Bisnis)	Atribut
<i>Class</i>	Kelas perjalanan di pesawat penumpang (Bisnis, Eco, Eco Plus)	Atribut
<i>Flight distance</i>	Jarak penerbangan	Atribut
<i>Inflight wifi service</i>	Tingkat kepuasan layanan wifi dalam pesawat (1-5)	Atribut
<i>Arrival time convenient</i>	Tingkat kepuasan waktu keberangkatan atau kedatangan (1-5)	Atribut
<i>Ease of Online booking</i>	Tingkat kepuasan pemesanan online (1-5)	Atribut
<i>Gate location</i>	Tingkat kepuasan lokasi Gerbang (1-5)	Atribut
<i>Food and drink</i>	Tingkat kepuasan makanan dan minuman (1-5)	Atribut
<i>Online boarding</i>	Tingkat kepuasan boarding online (1-5)	Atribut
<i>Seat comfort</i>	Tingkat kepuasan kenyamanan kursi (1-5)	Atribut
<i>Inflight entertainment</i>	Tingkat kepuasan hiburan dalam pesawat (1-5)	Atribut
<i>On-board service</i>	Tingkat kepuasan layanan <i>On-board</i> (1-5)	Atribut
<i>Leg room service</i>	Tingkat kepuasan layanan kamar kaki (1-5)	Atribut
<i>Baggage handling</i>	Tingkat kepuasan penanganan bagasi (1-5)	Atribut

<i>Check-in service</i>	Tingkat kepuasan layanan <i>Check-in</i> (1-5)	Atribut
<i>Inflight service</i>	Tingkat kepuasan layanan dalam pesawat (1-5)	Atribut
<i>Cleanliness</i>	Tingkat kepuasan kebersihan (1-5)	Atribut
<i>Departure Delay</i>	Menit ditunda saat keberangkatan	Atribut
<i>Arrival Delay</i>	Menit tertunda saat kedatangan	Atribut
<i>Satisfaction</i>	Tingkat kepuasan maskapai penerbangan (Puas, Tidak Puas/Netral)	Label

Data Training & Testing

Dataset yang didapatkan dari website kaggle sendiri sudah terbagi menjadi 2 bagian yaitu, data training dan data testing. Setelah memasukkan data training dan data testing ke dalam sebuah variable maka selanjutnya untuk memudahkan proses analisis data, data tersebut akan digabungkan terlebih dahulu dengan melakukan penambahan fitur yang bernama ‘code’ yang menandakan apakah data tersebut termasuk data training atau data testing. Berikut sample data dari dataset yang akan digunakan.

Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	4	3	4	4	5
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	5	3	1	4
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	4	3	4	4	4
3	3	24028	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	5	3	1	4
4	4	119299	Male	Loyal Customer	01	Business travel	Business	214	3	3	...	3	4	4	3	3

Gambar 1. Sampel Data

Seleksi Fitur

Adapun fitur yang tidak digunakan pada penelitian ini adalah *Unamed: 0, id, Gender dan Age*. Untuk *Unamed: 0* dan *id* sendiri merupakan nomor baris yang terdapat pada csv sebelumnya dan id dari setiap baris/record, kedua atribut ini tidak akan digunakan. Sedangkan untuk *gender* dan *age* karena kedua atribut ini tidak memiliki arti yang berarti terhadap klasifikasi kepuasan maskapai penerbangan.

Tipe Data

Pada analisis data dilakukan pengecekan tipe data dari setiap fitur dari *dataset* yang digunakan.

Berikut tipe data dari setiap fitur yang terdapat pada *dataset*.

Atribut/Label	Tipe Data
Unnamed: 0	Integer (Numeric)
id	Integer (Numeric)
Gender	Object (Categorical)
Customer Type	Object (Categorical)
Age	Integer (Numeric)
Type of Travel	Object (Categorical)
Class	Object (Categorical)
Flight Distance	Integer (Numeric)
Inflight wifi service	Integer (Numeric)
Departure/Arrival time convenient	Integer (Numeric)
Ease of Online booking	Integer (Numeric)
Gate location	Integer (Numeric)
Food and drink	Integer (Numeric)
Online boarding	Integer (Numeric)
Seat comfort	Integer (Numeric)
Inflight entertainment	Integer (Numeric)
On-board service	Integer (Numeric)
Leg room service	Integer (Numeric)
Baggage handling	Integer (Numeric)
Checkin service	Integer (Numeric)
Inflight service	Integer (Numeric)
Cleanliness	Integer (Numeric)
Departure Delay in Minutes	Integer (Numeric)
Arrival Delay in Minutes	Float (Numeric)
satisfaction	Object (Categorical)

Dari tabel diatas telah ditemukan bahwa terdapat beberapa fitur yang bertipe data kategorik, yaitu *Customer Type, Type of Travel dan Class*. Ketiga fitur ini harus diubah tipe datanya dari yang semula bersifat kategorik menjadi numerik yaitu dengan menggunakan salah satu teknik *features engineering* yang terdapat di *library pandas*, yaitu *get_dummies* dimana teknik tersebut serupa dengan teknik *one-hot encoding* yaitu dengan membuat tabel-tabel yang berisikan nilai *boolean* yaitu *true* (1) dan *false* (0). Dimana yang menjadi pembeda dari kedua teknik tersebut hanyalah jumlah tabel yang dihasilkan. Dengan menggunakan teknik *get_dummies* jumlah tabel yang dihasilkan lebih sedikit dibandingkan dengan menggunakan *one-hot*

encoding. Berikut sampel data dari hasil teknik *get_dummies* pada ketiga fitur tersebut.

Tabel 3. Hasil teknik *get_dummies* pada atribut kategorik

<i>Customer Type_disloyal Customer</i>	<i>Type of Travel_Personal Travel</i>	<i>Class_Eco</i>	<i>Class_Eco Plus</i>
0	1	0	1
1	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Identify Duplicate Rows

Dalam tahap ini dilakukan untuk mengetahui apakah terdapat baris data yang duplikat. Dan dari hasil pengecekan dengan *Python* ditemukan bahwa terdapat 7 baris data yang terdapat duplikasi. Ketujuh baris data ini akan dihapus agar *dataset* memiliki data yang banyak dan beragam.

Customer Type	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service
Loyal Customer	Business travel	Business	1725	4	4	4	4	2	3	4	4	4	3	4
Loyal Customer	Business travel	Business	480	4	4	4	4	2	5	4	4	4	3	4
Loyal Customer	Business travel	Business	300	5	5	5	5	4	4	4	4	4	3	4
Loyal Customer	Business travel	Business	1580	2	2	2	2	5	5	4	3	5	5	5
Loyal Customer	Business travel	Business	997	4	4	4	4	3	5	5	5	5	4	5
Loyal Customer	Business travel	Business	3550	5	5	5	5	2	5	4	4	4	5	4
Loyal Customer	Business travel	Business	2475	1	1	1	1	5	4	5	5	5	5	5

7 rows x 22 columns

Gambar 2. Data Duplikat

Identify Missing or Null Values

Pada tahap ini dilakukan pengecekan untuk mengetahui apakah terdapat baris data yang mempunyai nilai kosong atau *null*. Dan dari hasil pengecekan dengan *Python* ditemukan bahwa terdapat 393 baris data pada fitur *Arrival Delay in Minutes* yang mempunyai nilai kosong atau *null*. Ke-393 baris data ini akan diisi dengan nilai median dari fitur *Arrival Delay in Minutes*.

Outliers Detection

Pada tahap ini dilakukan pengecekan untuk mengetahui apakah terdapat fitur yang memiliki nilai yang berbeda pada umumnya. Hal ini dilakukan karena nilai *outlier* pada suatu fitur akan menyebabkan rusaknya nilai perhitungan mean, karena rata-rata yang muncul tidak dapat mewakili data keseluruhan. Berdasarkan hasil dari pengecekan dengan menggunakan *Phyton* terdapat 3 fitur yang memiliki nilai *outlier*, yaitu *Flight Distance*, *Departure Delay in Minutes* dan *Arrival*

Delay in Minutes. Berikut tabel yang menjelaskan secara singkat terhadap 3 fitur tersebut.

Tabel 4. Detail Fitur

	<i>Flight Distance</i>	<i>Departure Delay in Minutes</i>	<i>Arrival Delay in Minutes</i>
mean	1190.295057	14.714506	15.046276
std	997	38.071999	38.417229
min	31	0	0
25%	414	0	0
50%	844	0	0
75%	1744	12	13
max	4983	1592	1584

Normalization

Pada penelitian ini ditemukan bahwa terdapat fitur yang memiliki nilai *outlier*. Dimana nilai *outlier* ini akan menyebabkan rusaknya hasil perhitungan *mean*, karena rata-rata yang muncul tidak dapat mewakili data keseluruhan. Dan juga nilai *outlier* ini akan merusak hasil pemodelan karena nilainya tidak wajar. Sehingga harus dilakukan penanganan terhadap nilai *outlier* tersebut. Salah satu cara yang dapat menangani masalah tersebut adalah dengan melakukan salah satu teknik dari normalisasi data yaitu dengan standarisasi data. Dalam tahap normalisasi data ini, penulis menggunakan salah satu *library* yaitu *StandarScaler* dalam melakukan standarisasi data.

```
[[-0.73152929, 0.20363704, 0.61620409, ..., 1.49055134,
-0.904374, 3.58666391],
[-0.95717559, 0.20363704, -0.69520575, ..., -0.67089269,
-0.904374, -0.27881062],
[-0.04757027, -0.54948583, -0.69520575, ..., -0.67089269,
-0.904374, -0.27881062],
...,
[0.80787993, -1.30260871, -1.35091068, ..., -0.67089269,
-0.904374, -0.27881062],
[-0.18997816, -1.30260871, -1.35091068, ..., -0.67089269,
1.10573723, -0.27881062],
[0.53509862, -1.30260871, -0.03950083, ..., -0.67089269,
-0.904374, -0.27881062]]])
```

Gambar 3. Sampel Hasil Normalization

Model yang diusulkan

Tahapan ini melibatkan algoritma klasifikasi yang dibangun sesuai dengan tujuan awal penelitian. Pengujian *dataset* untuk mengklasifikasi label *neutral or dissatisfied* dan *satisfied* pada penelitian ini menggunakan empat algoritma *Decision Tree*, *Random Forest*, *SVM* dan *K-NN* yang diimplementasikan dengan menggunakan bahasa pemrograman *Python* dan *Jupyter Notebook*

sebagai IDE (*Integrated Development Environment*) nya. Dan di dalam setiap algoritma klasifikasi yang akan dibangun akan dilakukan pengecekan beberapa *hyperparameter* dengan menggunakan *Grid Search CV*. Hal ini dilakukan agar dapat mengetahui parameter yang tepat pada setiap model algoritma dan dapat memberikan hasil tingkat akurasi yang lebih tinggi.

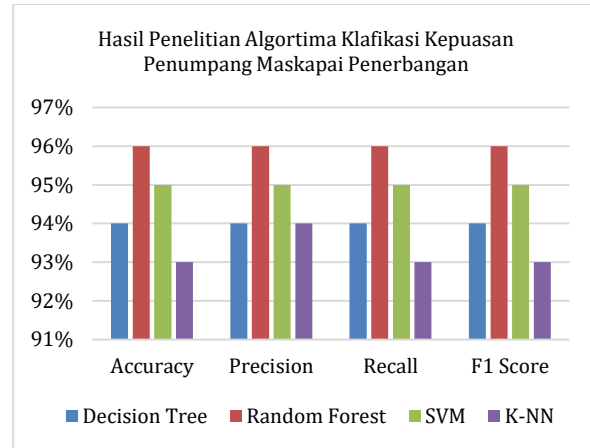
Hasil

Berdasarkan hasil penelitian yang dilakukan pada *dataset Airplane Passenger Satisfaction* dengan melakukan *split data*, salah satu teknik yang digunakan pada penelitian ini yaitu dengan melakukan evaluasi hasil akurasi dari setiap algoritma yang digunakan dengan menggunakan 80% dari *dataset* sebanyak 103.904 data sebagai data *training* dan 20% dari *dataset* sebanyak 25.976 data sebagai data *testing*. *Dataset* ini memiliki 22 atribut/fitur sebagai variable input dan '*Satisfaction*' sebagai label/target. Hasil penelitian akan dijelaskan serta dapat dilihat lebih rincinya pada tabel dan akan ditampilkan dalam bentuk grafik histogram, sebagai berikut ini:

Tabel 5. Tingkat akurasi algoritma yang digunakan

Classifier	Accuracy	Precision	Recall
Decision Tree	94%	94%	94%
Random Forest	96%	96%	96%
SVM	95%	95%	95%
K-NN	93%	94%	93%

Berdasarkan pada tabel diatas, hasil dari penelitian klasifikasi model hasil *Random Forest* merupakan hasil dengan akurasi tertinggi yaitu 96% dengan nilai *Precision* 96%, *Recall* 96% dan *F1 Score* 96%. Hasil nilai tertinggi kedua dengan selisih akurasi sekitar 1% yaitu model *SVM* dengan nilai *Precision* 95%, *Recall* 95% dan *F1 Score* 95%. Yang ketiga dengan selisih 2% yaitu *Decision Tree* dengan nilai *Precision* 94%, *Recall* 94% dan *F1 Score* 94%. Dan yang terakhir dengan selisih 3% yaitu *K-NN* dengan nilai *Precision* 94%, *Recall* 93% dan *F1 Score* 93%. Dari hasil uraian diatas dapat disimpulkan bahwa algoritma *Random Forest* merupakan algoritma klasifikasi dengan nilai tertinggi dan menjadi algoritma klasifikasi yang baik untuk *dataset Airplane Passenger Satisfaction*.



Gambar 4. Hasil Perbandingan 4 Algoritma

Berdasarkan pada grafik diatas dapat disimpulkan bahwa algoritma klasifikasi yang memiliki *accuracy*, *precision*, *recall* dan *f1-score* tertinggi yaitu *Random Forest* dengan nilai rata-rata 96%.

Confussion Matrix

Model *Confussion Matrix* akan membentuk matrix yang terdiri dari *true positive*, *false positive*, *true negative* dan *false negative*. Berikut dibawah ini merupakan hasil dari *confussion matrix* dari algoritma klasifikasi *Decision Tree*, *Random Forest*, *SVM* dan *K-NN*.

Hasil perhitungan akurasi untung algoritma *Decision Tree* dapat dilihat pada tabel dibawah ini.

Tabel 6. Tingkat akurasi algoritma yang digunakan

Actual/True Label	Predicted Label	
	neutral or dissatisfied	satisfied
Decision Tree	13779	794
	830	10572

Data label *neutral or dissatisfied* yang sesuai dengan prediksi yaitu sebanyak 13.779 *record*. Data *satisfied* yang termasuk ke dalam *neutral or dissatisfied* yaitu 794 *record*. Data *satisfied* yang sesuai dengan prediksi yaitu 10.572 *record*. Data *neutral or dissatisfied* yang termasuk ke dalam *satisfied* sebanyak 830 *record*.

Selanjutnya, hasil perhitungan akurasi untung algoritma *Random Forest* dapat dilihat pada tabel dibawah ini.

Tabel 7. Tingkat akurasi algoritma yang digunakan

<i>Confussion Matrix Random Forest</i>		<i>Predicted Label</i>	
		<i>neutral or dissatisfied</i>	<i>satisfied</i>
<i>Actual/True Label</i>	<i>neutral or dissatisfied</i>	14270	303
	<i>satisfied</i>	690	10712

Data label *neutral or dissatisfied* yang sesuai dengan prediksi yaitu sebanyak 14.270 *record*. Data *satisfied* yang termasuk ke dalam *neutral or dissatisfied* yaitu 303 *record*. Data *satisfied* yang sesuai dengan prediksi yaitu 10.712 *record*. Data *neutral or dissatisfied* yang termasuk ke dalam *satisfied* sebanyak 690 *record*.

Selanjutnya, hasil perhitungan akurasi untung algoritma *SVM* dapat dilihat pada tabel dibawah ini.

Tabel 8. Tingkat akurasi algoritma yang digunakan

<i>Confussion Matrix SVM</i>		<i>Predicted Label</i>	
		<i>neutral or dissatisfied</i>	<i>satisfied</i>
<i>Actual/True Label</i>	<i>neutral or dissatisfied</i>	14109	464
	<i>satisfied</i>	735	10667

Data label *neutral or dissatisfied* yang sesuai dengan prediksi yaitu sebanyak 14.109 *record*. Data *satisfied* yang termasuk ke dalam *neutral or dissatisfied* yaitu 464 *record*. Data *satisfied* yang sesuai dengan prediksi yaitu 10.667 *record*. Data *neutral or dissatisfied* yang termasuk ke dalam *satisfied* sebanyak 735 *record*.

Terakhir, hasil perhitungan akurasi untuk algoritma *K-NN* dapat dilihat pada tabel dibawah ini.

Tabel 9. Tingkat akurasi algoritma yang digunakan

<i>Confussion Matrix K-NN</i>		<i>Predicted Label</i>	
		<i>neutral or dissatisfied</i>	<i>satisfied</i>
<i>Actual/True Label</i>	<i>neutral or dissatisfied</i>	14151	422
	<i>satisfied</i>	1309	10000

Data label *neutral or dissatisfied* yang sesuai dengan prediksi yaitu sebanyak 14.151 *record*. Data *satisfied* yang termasuk ke dalam *neutral or dissatisfied* yaitu 422 *record*. Data *satisfied* yang sesuai dengan prediksi yaitu 10.000 *record*. Data *neutral or dissatisfied* yang termasuk ke dalam *satisfied* sebanyak 1309 *record*.

KESIMPULAN

Dalam menentukan indikator kepuasan penumpang pada studi kasus “Kepuasan Maskapai Penerbangan” dapat diprediksi salah satunya dengan menggunakan teknik *data mining*. Klasifikasi data merupakan salah satu penerapan dari *data mining*. Pada penelitian ini dalam melakukan penerapan klasifikasi data tersebut menggunakan algoritma *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)*. Dari hasil penelitian yang dilakukan dapat diketahui bahwa *Random Forest* merupakan algoritma klasifikasi yang paling baik dalam studi kasus ini, dengan nilai tingkat akurasi yaitu 96%. Adapun untuk *Decision Tree*, *Support Vector Machine (SVM)* dan *K-Nearest Neighbor* masing-masing memiliki tingkat akurasi 94%, 95% dan 93%. Sehingga pada studi kasus “Kepuasan Maskapai Penerbangan” *Random Forest* merupakan algoritma yang terbaik dalam melakukan prediksi dalam menentukan indikator kepuasan penumpang.

REFERENSI

- Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39–43. <https://doi.org/10.33096/ijodas.v1i2.13>
- . m, S., Marlita, D., & Wardana, A. (2018). Kualitas Layanan , Kepercayaan Merek dan Kepuasan Penumpang. *Jurnal Manajemen Transportasi & Logistik*, 05(02), 169–178
- Lestari, E. R. (2019). *Manajemen Inovasi: Upaya Meraih Keunggulan Kompetitif*. Universitas Brawijaya Press.
- Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(1), 27. <https://doi.org/10.24014/ijaidm.v1i1.4903>
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna

- Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>.
- Rudi, A., & Asrul. (2018). *Pengantar Sistem dan Perencanaan Transpotasi*. Deepublish.
- Siti Fatimah. (2019). Pengantar Transportasi. In *Myria Publisher*. Myria Publisher.
- Widjaja, E. L. (2017). Analisa Pengaruh Kualitas Layanan Terhadap Kepuasan Penumpang Maskapai Penerbangan Batik Air. *Universitas Kristen Petra, 2013(Lcc)*, 224–233.
- Wijayanto, A., Bernardo, J. F. A., & Pamungkas, S. (2021). Analisis Klasifikasi Kepuasan Penumpang Maskapai Penerbangan Menggunakan Algoritma Naïve Bayes. *Jurnal Sains Komputer Dan Teknologi Informasi*, 3(2), 97–103
- Yoga Religia, & Amali. (2021). Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 527–5
- Yuliana, Paradise, & Kusri. (2021). Sistem Pakar Diagnosa Penyakit Ispa Menggunakan Metode Naive Bayes Classifier Berbasis Web. *CSRID (Computer Science Research and Its Development Journal)*, 10(3), 127. <https://doi.org/10.22303/csrid.10.3.2018.127-138>