

UJI EFEKTIVITAS PENERAPAN *MACHINE LEARNING CLASSIFICATION* UNTUK *SURVEY* KEPUASAN PELANGGAN MASKAPAI PENERBANGAN X

EFFECTIVENESS TEST OF APPLICATION OF MACHINE LEARNING CLASSIFICATION FOR AIRLINE CUSTOMER SATISFACTION SURVEY X

¹Dimas Aditiya*, ²Ulinnuha Latifa

^{1,2}Program Studi Teknik Elektro, Fakultas Teknik, Universitas Singaperbangsa Karawang

¹dimas.aditiya18145@student.unsika.ac.id, ²ulinnuha.latifa@ft.unsika.ac.id

INFO ARTIKEL

Diterima: 18 Mei 2022

Direvisi: 27 Juni 2022

Disetujui: 08 Juli 2022

Kata Kunci:

Cross Validation, Decision Tree, Machine Learning, Maskapai Penerbangan, Random Forest, Supervised Learning.

Keywords:

Airlines, Cross Validation, Decision Tree, Machine Learning, Random Forest, Supervised Learning.

ABSTRAK

Transportasi dapat mempermudah masyarakat untuk berpindah dari satu tempat ke tempat lainnya secara cepat, salah satunya dengan menggunakan transportasi udara. Karena hal ini maskapai penerbangan perlu meningkatkan kualitas pelayanan melalui survey yang diberikan kepada penumpang. Proses klasifikasi kepuasan pelanggan melalui survey maskapai penerbangan dapat dilakukan dengan mudah jika menggunakan metode *data mining*. Penelitian ini berfokus pada proses klasifikasi kepuasan pelanggan dibuat menggunakan *machine learning* dengan model *K-nearest neighbor*, *decision tree*, dan *random forest* agar dapat mempermudah maskapai dalam melakukan klasifikasi. Sedangkan performa model *machine learning* klasifikasi dianalisis berdasarkan hasil akurasi, presisi, *recall*, dan *F1-Score*. Berdasarkan hasil Analisis performa yang dilakukan, model *machine learning decision tree* dan *random forest* memiliki hasil performa yang sangat baik dengan nilai akurasi *decision tree* sebesar 92,96% pada data *testing* dan *random forest* memiliki nilai akurasi sebesar 93,22% pada data *testing*. Kedua model *machine learning* tersebut juga diuji menggunakan metode *cross validation* untuk mengetahui model *machine learning* yang lebih efektif untuk diterapkan. Berdasarkan hasil uji *cross validation* model *decision tree* memiliki akurasi sebesar 94,5% dan *random forest* memiliki akurasi sebesar 96%. Dengan nilai *cross validation* yang bagus, maka *decision tree* dan *random forest* dapat diterapkan untuk membantu maskapai penerbangan X dalam menentukan nilai kepuasan pelanggan.

ABSTRACT

The use of air transportation is one way that transportation may make it simpler for people to get from one place to another place rapidly. As a result, airlines must enhance the quality of their services by using passengers feedback. Using the data mining method makes it simple to categorize consumer satisfaction from airline surveys. This study focuses on the customer satisfaction classification method created using machine learning with the K-nearest neighbor, decision tree, and random forest models to make it simpler for airlines to categorize. The accuracy, precision, recall, and F1-Score statistics are used to analyze the performance of the classification machine learning model. According to the findings of the performance study, the machine learning decision tree and random forest models have good performance results. The accuracy values for the testing data for the decision tree and random forest models, respectively, are 92,96% and 93,22%. The cross-validation method was also used to examine the two machine learning models to determine which one is more practical to use. The decision tree model and the random forest both have accuracy levels of 96% and 94,5%, respectively, according to the findings of the cross-validation test. Decision trees and random forests can be used to help airline X determine customer satisfaction levels if the cross-validation value is increased.

*Corresponding author: dimas.aditiya18145@student.unsika.ac.id

I. PENDAHULUAN

Terjadinya peningkatan mobilitas dalam kehidupan manusia membawa dampak positif bagi dunia transportasi, baik itu transportasi darat, laut, dan udara. Transportasi dapat mempermudah masyarakat untuk berpindah dari satu tempat ke tempat lainnya secara cepat, salah satunya dengan menggunakan transportasi udara. Tingginya minat masyarakat untuk berpergian menggunakan transportasi udara dinilai sebagai potensi yang besar bagi maskapai penerbangan untuk meraup keuntungan melalui penyediaan

jasa transportasi udara [1]. Agar mendapatkan laba yang maksimal, maskapai penerbangan perlu meningkatkan kualitas pelayanan yang disediakan agar dapat membuat para pelanggan maskapai tersebut tidak segan untuk kembali menggunakan jasa layanan penerbangan tersebut.

Peningkatan kualitas pelayanan suatu maskapai perlu dilakukan agar pelanggan merasa puas terhadap pelayanan yang diberikan, pada penelitian sebelumnya yang dilakukan oleh Hayadi (2021), dapat diketahui bahwa ketika dalam penerbangan, kebanyakan pelanggan menilai tingkat

kepuasan mereka melalui seberapa bagus pelayanan maskapai dalam penerbangan mulai dari pelayanan pramugari hingga ketersediaan fasilitas pesawat yang memadai karena berkaitan langsung dengan konsumen [2]. Peningkatan kualitas pelayanan dalam penerbangan dapat mendorong keberhasilan sebuah maskapai penerbangan dalam meningkatkan jumlah kepuasan pelanggan, khususnya meningkatkan layanan makanan dalam penerbangan merupakan salah satu penentu yang signifikan dari layanan dalam penerbangan. Untuk mengetahui tingkat kepuasan pelanggan terhadap pelayanan yang diberikan, maskapai penerbangan perlu untuk melakukan survey agar dapat mengetahui layanan dan fasilitas apa saja yang dapat membuat pelanggannya merasa puas. Survey dilakukan agar pihak maskapai penerbangan bisa mendapatkan informasi masukan terhadap pelayanan yang diberikan agar bisa disesuaikan dengan keinginan pelanggan dan menjadi bahan evaluasi untuk meningkatkan fasilitas pelayanan yang dirasa performanya kurang baik.

Analisis data survey untuk mengetahui tingkat kepuasan pelanggan maskapai penerbangan akan menjadi sebuah permasalahan baru jika analisis yang dilakukan masih secara manual dan data yang harus dianalisis berjumlah ratusan ribu atau bahkan jutaan pelanggan perharinya, maka salah satu cara untuk mengatasi permasalahan tersebut adalah dengan memanfaatkan perkembangan teknologi dan era *big data* yang sedang berkembang saat ini. *Big data* merupakan sebuah istilah untuk menggambarkan data yang memiliki tingkat keberagaman yang banyak, serta memiliki volume dan kecepatan yang tinggi dalam penyimpanan informasinya [3]. *Big data* memerlukan metode dan perangkat yang sesuai agar dapat melakukan pengolahan data untuk menciptakan berbagai informasi penting, pengambilan keputusan, serta melakukan otomatisasi proses.

Proses klasifikasi kepuasan pelanggan maskapai penerbangan dapat dilakukan dengan mudah jika menggunakan metode *data mining*. *Data mining* adalah sebuah proses pengambilan keputusan yang tepat dengan melakukan pencarian pola data maupun informasi *raw data* menggunakan algoritma *machine learning* [4]. *Data mining* dapat digunakan untuk mengevaluasi model transendental yang dapat bertindak sebagai jembatan antara teori dan praktik berdasarkan kumpulan data yang besar [5].

Dalam berbagai kasus yang dihadapi oleh keterbatasan manusia akan kemampuannya, penggunaan *machine learning* dapat membantu manusia untuk melakukan pekerjaan yang kompleks menjadi mudah dan cepat untuk diselesaikan. *Machine learning* merupakan sebuah metode populer yang digunakan untuk menggantikan atau menirukan perilaku manusia dalam menyelesaikan masalah dan melakukan otomatisasi [6]. Umumnya *machine learning* dapat diterapkan untuk melakukan klasifikasi dan prediksi, karena *machine learning* bekerja dengan cara memahami pola historis data yang terjadi. Pada dasarnya *machine learning* memiliki tiga buah cara kerja yang biasa disebut dengan metode *supervised learning*, *unsupervised learning* dan *reinforcement learning* [7].

Supervised learning merupakan cara kerja *machine learning* dengan menggunakan metode klasifikasi dimana kumpulan data sepenuhnya diberikan label untuk mengklasifikasikan kelas yang tidak dikenal. Sedangkan *unsupervised learning* merupakan cara kerja *machine learning* dengan menggunakan metode *clustering* dikarenakan tidak adanya kebutuhan untuk memberi label pada kumpulan data yang ada dan hasilnya tidak untuk mengidentifikasi contoh kelas yang telah ditentukan sebelumnya [8]. Sedangkan *reinforcement learning* memiliki metode yang berada diantara *supervised learning* dan *unsupervised learning*, cara kerja dari metode ini adalah dengan menyelesaikan tujuan sesuai dengan konsepnya tanpa ada pemberitahuan secara eksplisit dari komputer jika tujuan tersebut telah tercapai [9]. Contoh penggunaan *reinforcement learning* adalah untuk sistem rekomendasi yang biasa digunakan oleh berbagai perusahaan besar seperti Youtube, Netflix, dan Spotify.

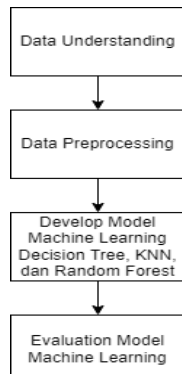
Penelitian yang digunakan untuk melakukan uji efektivitas *machine learning* klasifikasi kali ini dilakukan dengan menggunakan cara kerja *supervised learning*, dimana target akhir dari kumpulan data yang dimiliki berupa nilai kepuasan dari hasil survey pelanggan maskapai penerbangan dengan nilai *satisfied* atau *neutral/dissatisfied*. Penelitian bertujuan membuktikan apakah *machine learning* dapat bekerja secara efektif dalam mengklasifikasi nilai kepuasan pelanggan penerbangan. Penentuan nilai kepuasan hasil survey ditentukan dengan analisa klasifikasi. Analisa klasifikasi merupakan suatu metode untuk menemukan model *machine learning* yang dapat menjelaskan proses klasifikasi data yang digunakan untuk memprediksi kelas, label, maupun sebuah kategori dari sebuah *dataset* [10]. *Machine learning* klasifikasi memiliki beberapa algoritma, diantaranya *K-Nearest Neighbor*, *Decision Tree*, *Support Vector Machines*, *Recurrent Neural Network (RNN)*, *Random Forest*, dan *Convolutional Neural Network (CNN)*. *Machine learning* RNN dan CNN merupakan jenis arsitektur Jaringan Syaraf Tiruan (JST), JST merupakan sebuah metode proses informasi dengan karakteristik seperti jaringan syaraf Biologi [11]. Pada penelitian ini untuk mengklasifikasikan dan memprediksi kategori kepuasan pelanggan berdasarkan hasil survey, digunakan algoritma *machine learning* jenis *Decision Tree*, *K-Nearest Neighbor*, dan *Random Forest*. Penggunaan model *machine learning* dapat membantu maskapai penerbangan untuk meningkatkan layanan penerbangan secara efisien berdasarkan hasil survey yang dimiliki agar dapat meningkatkan profit untuk perusahaan tersebut [12].

II. METODE PENELITIAN

Penelitian dilakukan dengan menggunakan platform google colab sebagai *software* untuk melakukan analisis *dataset* dan membangun model *machine learning* yang dibutuhkan, dengan tahapan metodologi penelitian yang terdiri dari 4 tahap seperti yang dapat dilihat dalam Gambar 1.

A. Data Understanding

Pada tahap *data understanding* merupakan tahap dilakukannya pengumpulan data, dan menganalisa data untuk memahami data yang digunakan. Penelitian ini menggunakan data yang diambil dari sebuah situs penyedia data milik google yaitu www.kaggle.com yang diakses pada tanggal 07 Oktober 2021 berupa data *Airline Passenger Satisfaction* [13].



Gambar 1 Metodologi penelitian

TABEL I
DESKRIPSI KOLOM

Nama Kolom	Deskripsi
ID	ID Penumpang
Gender	Jenis Kelamin Penumpang
Customer Type	Jenis Pelanggan
Age	Umur Penumpang
Type of Travel	Jenis Penerbangan Penumpang
Class	Kelas Perjalanan Penumpang
Flight Distance	Jarak Penerbangan Penumpang
Inflight Wifi Service	Tingkat Kepuasan Wifi (1-5)
Departure/Arrival Time Convenient	Tingkat Kepuasan Waktu Kedatangan/ Keberangkatan nyaman (1-5)
Ease of Online Booking	Tingkat Kepuasan Online Booking (1-5)
Gate Location	Tingkat Kepuasan lokasi gerbang (1-5)
Food and Drink	Tingkat Kepuasan Makanan dan Minuman (1-5)
Online Boarding	Tingkat Kepuasan Boarding Online (1-5)
Seat Comfort	Tingkat Kepuasan kenyamanan kursi (1-5)
Inflight Entertainment	Tingkat Kepuasan Hiburan dalam Pesawat (1-5)
On-board Service	Tingkat Kepuasan Layanan On-board (1-5)
Leg Room Service	Tingkat Kepuasan Ruang Kaki (1-5)
Baggage Handling	Tingkat Kepuasan Penanganan Bagasi (1-5)
Check-in Service	Tingkat Kepuasan Layanan Check-in (1-5)
Inflight Service	Tingkat Kepuasan Layanan Dalam Pesawat (1-5)
Cleanliness	Tingkat Kepuasan Kebersihan (1-5)
Departure Delay in Minutes	Keterlambatan Keberangkatan
Arrival Delay in Minutes	Keterlambatan Kedatangan
Satisfaction	Tingkat Kepuasan

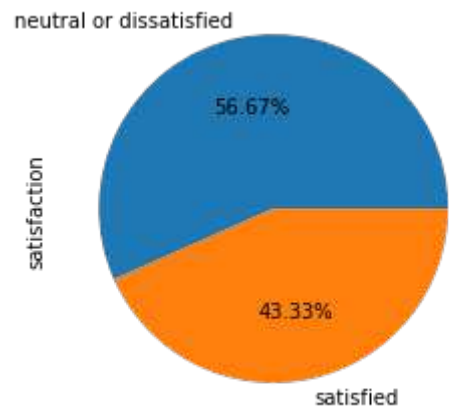
Pada Tabel I menunjukkan kolom yang terdapat pada data *Airline Passenger Satisfaction* beserta deskripsi dari masing-masing kolom tersebut. Pemilihan data *Airline Passenger Satisfaction* didasari karena pada data tersebut memiliki jumlah data survey yang banyak, yaitu sebanyak 103904 baris dan 24 kolom. Pada data *Airline Passenger Satisfaction* memiliki jumlah data yang seimbang sehingga tidak perlu dilakukan proses untuk mengatasi data yang *imbalanced*. Dikarenakan data survey yang dimiliki jumlahnya seimbang, maka data survey sudah dirasa cukup untuk dilakukan proses *sampling* guna mencari tahu efektivitas *machine learning* dalam klasifikasi nilai kepuasan pelanggan maskapai penerbangan x.

```

neutral or dissatisfied 58879
satisfied                45025
Name: satisfaction, dtype: int64
    
```

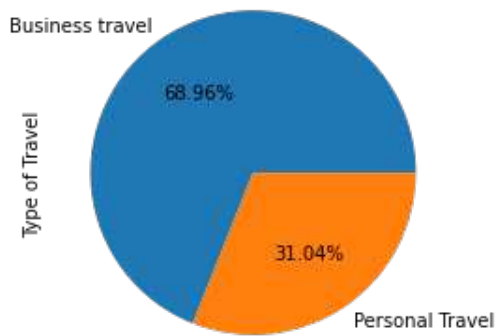
Gambar 2 Total jumlah penumpang puas dan tidak puas

Gambar 2 menunjukkan total jumlah penumpang yang merasa puas dan merasa tidak puas berdasarkan hasil survey yang telah dilakukan oleh maskapai, berdasarkan survey tersebut dapat diketahui bahwa sebanyak 58879 penumpang merasa netral/tidak puas dengan pelayanan yang diberikan. Sedangkan sebanyak 45025 penumpang lainnya merasa puas terhadap pelayanan yang diberikan.



Gambar 3 Presentase jumlah penumpang puas dan tidak puas

Pada Gambar 3 menunjukkan bahwa presentase penumpang yang merasa puas sebesar 43.33% dan yang merasa tidak puas dengan pelayanan yang diberikan maskapai penerbangan sebesar 56.67%. Melalui data tersebut dapat diketahui bahwa dataset *Airline Passenger Satisfaction* ini memiliki jumlah yang seimbang sehingga tidak perlu lagi dilakukan proses *resampling techniques* untuk mengatasi data yang *imbalanced*.

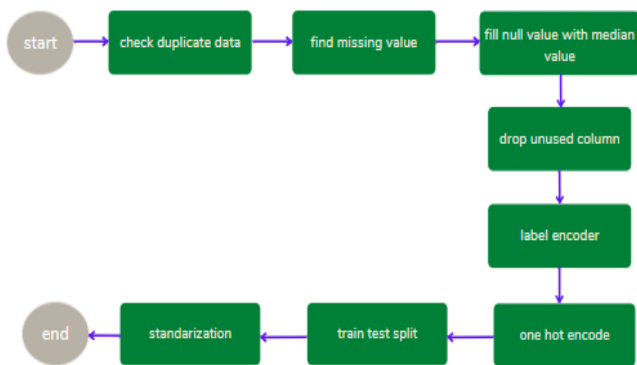


Gambar 4 Persentase tipe perjalanan penumpang

Melalui Gambar 4 dapat diketahui bahwa jumlah penumpang yang menggunakan maskapai penerbangan lebih banyak menggunakannya untuk perjalanan bisnis dibanding untuk perjalanan pribadi.

B. Data Preprocessing

Tahap ini merupakan proses yang dilakukan sebelum memasuki tahap pembuatan *machine learning*, *data preprocessing* merupakan tahapan untuk mengubah data mentah yang dikumpulkan menjadi sebuah informasi yang lebih bersih dan dapat digunakan untuk tahap pengolahan berikutnya. *Data preprocessing* merupakan sebuah proses untuk mengubah data mentah menjadi format data yang berguna dan efisien [14].



Gambar 5 Diagram alir *data preprocessing*

Jika dilihat pada Gambar 5, tahapan yang dilakukan untuk melakukan *data preprocessing* pada penelitian ini akan melalui delapan tahapan sebelum masuk ke tahap pembuatan model *machine learning*.

C. Develop Model Machine Learning

Untuk mengetahui model *machine learning* yang paling efektif untuk dataset survey kepuasan pelanggan maskapai penerbangan, maka setelah melalui tahap *data preprocessing*, selanjutnya memasuki tahap pembuatan model *machine learning* klasifikasi dengan menggunakan model *K-Nearest Neighbor (KNN)*, *Decision Tree*, dan *Random Forest*.

1. K-Nearest Neighbor (KNN)

Model *K-Nearest Neighbor* merupakan sebuah model *machine learning* klasifikasi yang menggunakan metode pembelajaran *dataset* berdasarkan jarak data pembelajaran terdekat. Jarak pembelajaran yang dimaksud adalah jarak *Euclidean* suatu data pembelajaran baik itu jarak yang jauh maupun dekat [15].

2. Decision Tree

Model *Decision Tree* merupakan sebuah model *machine learning* yang memiliki cara kerja seperti struktur menyerupai diagram alir yang setiap node merupakan tahap pengujian terhadap variabel atribut. *Decision Tree* digunakan untuk analisa klasifikasi dan prediksi dalam sebuah bentuk pohon keputusan [16].

3. Random Forest

Model *Random Forest* merupakan sebuah model *machine learning* yang cara kerjanya memiliki kemiripan dengan *Decision Tree*. Cara kerja *Random Forest* merupakan *base classifier* yang terdiri dari kumpulan pohon keputusan yang dikombinasikan. *Random Forest* biasanya digunakan untuk analisa klasifikasi dan regresi [17].

D. Evaluation Model Machine Learning

Pada tahap ini model *machine learning* yang telah diterapkan untuk melakukan klasifikasi akan dilakukan perhitungan performa model *machine learning* dengan menggunakan perhitungan yang didasarkan pada pengujian objek yang benar dan objek yang salah. Setiap model *machine learning* yang digunakan akan menghasilkan nilai performa berupa nilai akurasi, presisi, *recall* dan *F1-score* yang kemudian akan dibandingkan untuk dilakukan Analisa agar mengetahui model *machine learning* mana yang memiliki performa paling baik. Nilai evaluasi model *machine learning* yang berupa nilai akurasi, presisi, *recall*, dan *F1-Score* akan dihitung menggunakan *binary confusion matrix*.

Binary confusion terdiri dari hasil pengamatan yang diklasifikasikan dengan benar kedalam kelas negatif yang biasa disebut dengan *true negatives (TN)* dan pengamatan yang diklasifikasikan dengan benar kedalam kelas positif yang biasa disebut dengan *true positives (TP)*. Data dari kelas negatif yang diklasifikasikan salah sebagai positif biasa disebut dengan *false positives (FP)*, sedangkan data dari kelas positif yang diklasifikasikan salah sebagai negatif biasa disebut dengan *false negatives (FN)*. Kinerja pengklasifikasian *machine learning* dapat dihitung menggunakan performa klasifikasi yaitu TP, FP, TN, dan FN [18].

Pada penelitian ini performa model *machine learning* klasifikasi akan dianalisa berdasarkan hasil akurasi, presisi, *recall*, dan *F1-Score* yang dihitung berdasarkan persamaan (1), (2), (3), dan (4) berturut-turut menggunakan rumus perhitungan akurasi, presisi, *recall*, dan *F1-Score* [19].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Presisi = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

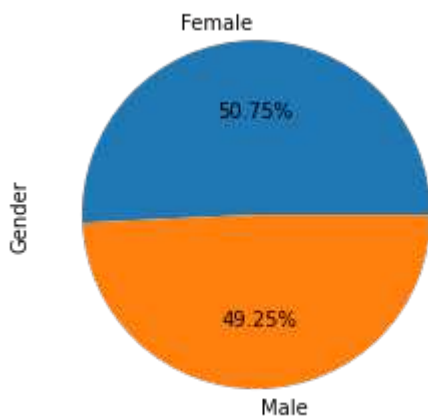
$$F1\ Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \tag{4}$$

Selain menggunakan *confusion matrix* sebagai hasil evaluasi performa model *machine learning*, pada penelitian ini juga menggunakan *validation technique* untuk mengukur seberapa optimal model *machine learning* yang telah dibuat. *Validation technique* dapat dilakukan dengan menggunakan dua cara, pertama dengan menggunakan teknik *cross validation* dan yang kedua dengan menggunakan teknik *holdout*. Pada penelitian ini berfokus untuk menggunakan *validation technique* dengan metode *cross validation*.

III. HASIL DAN PEMBAHASAN

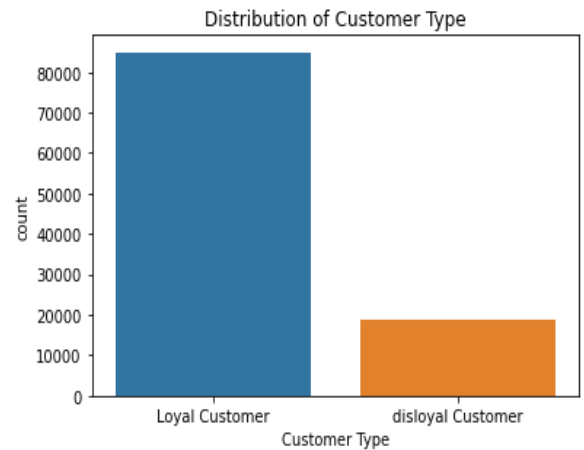
A. Data Understanding

Tahap *Data Understanding* merupakan tahap untuk melakukan pemahaman terhadap data yang kita miliki. Pada tahap ini juga dilakukan *exploratory data analysis*, *exploratory data analysis* merupakan sebuah metode untuk melakukan pemahaman data, Analisa data, serta menggali informasi penting yang bisa didapatkan dengan menggunakan bantuan visualisasi data ataupun statistic. Melalui metode *exploratory data analysis* pada project dataset *Airline Passenger Satisfaction*, berhasil ditemukan beberapa informasi penting yang akan berguna untuk kepentingan analisa.



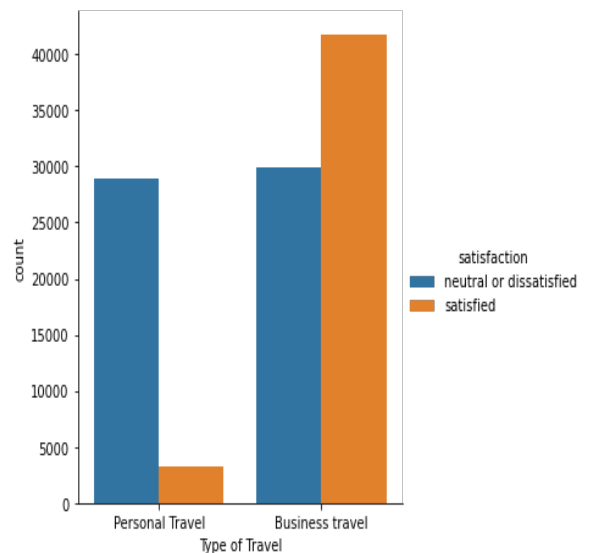
Gambar 6 Persentase jumlah penumpang berdasarkan gender

Gambar 6 menunjukkan bahwa persentase penumpang Pria sebesar 49.25% dan penumpang wanita sebesar 50.75%. Hal ini menunjukkan distribusi data antara penumpang Pria dan Wanita jumlahnya tidak berbeda jauh.



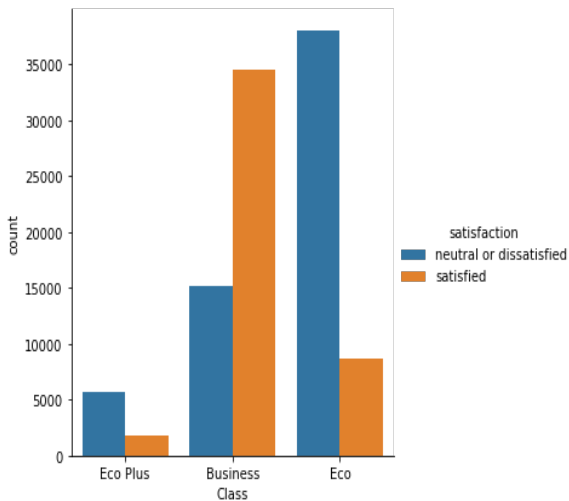
Gambar 7 Distribusi data customer type

Pada Gambar 7 menunjukkan bahwa terdapat perbedaan jumlah penumpang yang cukup jauh pada tipe penumpang yang menaiki maskapai penerbangan tersebut. Dapat dilihat bahwa *Loyal Customer* merupakan penumpang yang paling banyak. Maka dari itu maskapai penerbangan harus meningkatkan kualitas pelayanan mereka agar dapat menciptakan jumlah *loyal customer* yang semakin banyak karena merasa puas dengan pelayanan yang telah diberikan oleh maskapai penerbangan.



Gambar 8 Nilai kepuasan terhadap jenis perjalanan

Melalui Gambar 8 dapat diketahui bahwa para pelanggan yang menggunakan jasa maskapai penerbangan untuk perjalanan bisnis jauh merasa lebih puas dibandingkan dengan jenis perjalanan pribadi. Hal ini bisa menjadi bahan evaluasi untuk maskapai penerbangan agar dapat meningkatkan kualitas pelayanan untuk jenis perjalanan pribadi agar dapat meningkatkan nilai kepuasan mereka.



Gambar 9 Nilai kepuasan terhadap jenis kelas penerbangan

Pada Gambar 9 dapat diketahui bahwa penumpang yang menggunakan jenis kelas penerbangan bisnis jauh merasa lebih puas dengan pelayanan yang diberikan jika dibandingkan dengan jenis kelas penerbangan ekonomi yang dominan merasa tidak puas dengan pelayanan yang diberikan oleh maskapai penerbangan.

Setelah melakukan *exploratory data analysis*, pada penelitian ini berhasil mengelompokkan layanan maskapai penerbangan yang terbilang sudah cukup baik dan layanan yang perlu ditingkatkan lagi kualitasnya berdasarkan nilai survei kepuasan pelanggan. Layanan maskapai yang terbilang sudah cukup baik, diantaranya:

- Gate location
- Food and Drink
- Seat comfort
- Inflight entertainment
- On-board service
- Leg room service
- Checkin service
- Inflight service
- Cleanliness

Beberapa layanan tersebut merupakan layanan yang memiliki nilai kepuasan cukup baik dari pelanggan, sehingga layanan tersebut harus dijaga kualitas pelayanannya atau bahkan ditingkatkan lagi agar menjadi semakin baik. Sedangkan layanan maskapai penerbangan yang perlu ditingkatkan lagi kualitas pelayanannya adalah:

- *Baggage Handling* dengan nilai nol sebesar 6,97% meskipun banyak penumpang memberi rating 4.
- *Inflight wifi service* dengan nilai penilaian nol sebesar 2,99%.
- *Departure/Arrival time convenient* dengan nilai nol sebesar 5,10% meskipun banyak penumpang yang memberi rating 4.
- *Ease of online booking* dengan nilai nol sebesar 4,32%.
- *Online boarding* dengan nilai nol sebesar 2,34% meskipun banyak penumpang memberikan rating 4.

Beberapa layanan tersebut harus ditingkatkan lagi kualitasnya oleh maskapai penerbangan, karena peningkatan layanan merupakan salah satu faktor yang dapat mendatangkan untung lebih banyak kepada perusahaan dan dapat meningkatkan bisnisnya menjadi lebih baik.

B. Data Preprocessing

Tahap ini merupakan salah satu tahap penting yang harus dilakukan untuk membersihkan data agar dapat diciptakannya model *machine learning* dengan performa yang optimal. Setelah melakukan analisa untuk mengetahui adanya kemungkinan data yang duplikat pada *dataset Airline Passenger Satisfaction*, ternyata hasilnya nihil atau tidak ditemukan data yang duplikat pada *dataset* yang digunakan.

Gambar 8 menunjukkan bahwa terdapat sebanyak 310 nilai data yang *null* pada variabel *Arrival Delay in Minutes*, hal ini harus diatasi dengan pengisian nilai *null* tersebut agar tidak kosong.

Pada Gambar 9 menunjukkan hasil setelah nilai kosong atau nilai *null* diisi menggunakan nilai median. Nilai *null* diisi dengan nilai median dikarenakan distribusi pada variabel *Arrival Delay in Minutes* berbentuk normal.

Dalam proses pembuatan model *machine learning*, tidak semua variabel yang terdapat pada *dataset Airline Passenger Satisfaction* akan digunakan. Oleh karena itu, perlu dilakukan penghapusan terhadap variabel-variabel yang dinilai tidak akan berguna saat pembuatan model *machine learning*.

```

id 0
Gender 0
Customer Type 0
Age 0
Type of Travel 0
Class 0
Flight Distance 0
Inflight wifi service 0
Departure/Arrival time convenient 0
Ease of Online booking 0
Gate location 0
Food and drink 0
Online boarding 0
Seat comfort 0
Inflight entertainment 0
On-board service 0
Leg room service 0
Baggage handling 0
Checkin service 0
Inflight service 0
Cleanliness 0
Departure Delay in Minutes 0
Arrival Delay in Minutes 310
satisfaction 0
dtype: int64
    
```

Gambar 8 Menemukan nilai yang *null*

id	0
Gender	0
Customer Type	0
Age	0
Type of Travel	0
Class	0
Flight Distance	0
Inflight wifi service	0
Departure/Arrival time convenient	0
Ease of Online booking	0
Gate location	0
Food and drink	0
Online boarding	0
Seat comfort	0
Inflight entertainment	0
On-board service	0
Leg room service	0
Baggage handling	0
Checkin service	0
Inflight service	0
Cleanliness	0
Departure Delay in Minutes	0
Arrival Delay in Minutes	0
satisfaction	0

Gambar 9 Setelah mengisi nilai yang null

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking
0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	3
1	Male	disloyal Customer	25	Business travel	Business	235	3	2	3
2	Female	Loyal Customer	26	Business travel	Business	1142	2	2	2
3	Female	Loyal Customer	25	Business travel	Business	562	2	5	5
4	Male	Loyal Customer	61	Business travel	Business	214	3	3	3

Gambar 10 Penghapusan variabel yang tidak berguna

Pada *dataset Airline Passenger Satisfaction* terdapat dua variabel kolom yang tidak berguna yaitu kolom 'id' dan 'unnamed: 0'. Kedua kolom tersebut harus dihapus karena tidak butuhkan untuk membuat model *machine learning*. Gambar 10 menunjukkan beberapa variabel yang masih digunakan untuk proses pembuatan model *machine learning*.

	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink
0	0	13	1	2	460	3	4	3	1	5
1	1	25	0	0	235	3	2	3	3	1
2	0	26	0	0	1142	2	2	2	2	5
3	0	25	0	0	562	2	5	5	5	2
4	0	61	0	0	214	3	3	3	3	4

Gambar 11 Hasil proses *label encoder* dan *one hot encoder*

Machine learning tidak dapat membaca perintah dengan tipe data yang berbentuk huruf atau string, maka dari itu dilakukan pengubahan bentuk data dari yang berbentuk string menjadi angka seperti yang dapat dilihat pada Gambar 11 dengan menggunakan *label encoder* dan *one hot encode* agar *machine learning* dapat membaca isi dari variabel tersebut.

	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient
0	2.115103	-0.952878	-0.669448	-0.954884	-0.600675	0.956316	0.617626
1	-0.472790	-2.144383	-0.669448	-0.954884	0.247906	1.709937	1.273866
2	-0.472790	-0.555710	1.493767	0.655896	-0.266659	-0.550925	-0.038613
3	-0.472790	1.562521	1.493767	0.655896	0.309092	0.956316	1.273866
4	2.115103	-1.283852	-0.669448	0.655896	-0.724050	0.956316	-0.038613

Gambar 12 Hasil proses standardisasi dan *train test split*

Pada *dataset Airline Passenger Satisfaction* memiliki variabel *Flight Distance* sehingga perlu dilakukan standardisasi pada variabel tersebut agar nilai dari masing-masing jarak tempuh penerbangan bisa memiliki batas tertingginya seperti yang dilakukan pada Gambar 12. Sedangkan untuk membagi *dataset training* dan *dataset testing* digunakan metode *train test split* dengan persentase 80% untuk data *training* dan 20% untuk data *testing*.

C. Develop Model Machine Learning

Setelah selesai melakukan *Data Preprocessing*, langkah selanjutnya adalah membangun model *machine learning*, model *machine learning* yang dibutuhkan adalah model yang dapat mengklasifikasikan apakah penumpang merasa puas terhadap pelayanan maskapai atau tidak berdasarkan hasil survei yang telah dilakukan pihak maskapai. Dalam penelitian ini ada tiga buah model klasifikasi *machine learning* yang dibuat, diantaranya ada *K-Nearest Neighbor*, *Decision Tree*, dan *Random Forest*.

1. K-Nearest Neighbor (KNN)

Pada penelitian ini dalam proses pembuatan model *K-Nearest Neighbor* menggunakan nilai $K = 7$ dengan jarak *Euclidean* yang telah dinormalisasi.

	precision	recall	f1-score	support
dissatisfied	0.93	0.98	0.95	47036
satisfied	0.97	0.90	0.93	36087
accuracy			0.94	83123
macro avg	0.95	0.94	0.94	83123
weighted avg	0.94	0.94	0.94	83123

Gambar 13 Hasil evaluasi performa model KNN pada data *training*

Pada Gambar 13 dapat diketahui bahwa hasil evaluasi metrik pada data *training* model *machine learning K-Nearest Neighbor* memiliki nilai presisi *satisfied* sebesar 97% dan *dissatisfied* sebesar 93%. Nilai akurasi pada data

training model *machine learning* KNN juga dinilai sudah cukup bagus karena nilai akurasinya sebesar 94%.

	precision	recall	f1-score	support
dissatisfied	0.91	0.97	0.94	11843
satisfied	0.95	0.87	0.91	8938
accuracy			0.93	20781
macro avg	0.93	0.92	0.92	20781
weighted avg	0.93	0.93	0.93	20781

Gambar 14 Hasil evaluasi performa model KNN pada data *testing*

Pada Gambar 14 dapat diketahui bahwa hasil evaluasi metrik pada data *testing* model *machine learning* *K-Nearest Neighbor* memiliki nilai presisi *satisfied* sebesar 95% dan *dissatisfied* sebesar 91%. Nilai akurasi pada data *testing* model *machine learning* KNN juga dinilai memiliki performa yang baik karena nilai akurasinya sebesar 93%. Hal ini menunjukkan bahwa model *machine learning* yang dibuat sudah dapat melakukan pembelajaran dengan baik sehingga hasil performa model yang dihasilkan tidak *underfitting* ataupun *overfitting*.

2. Decision Tree

Pada penelitian ini dalam proses pembuatan model *Decision Tree* menggunakan nilai *max_depth* sebesar 15 yang berfungsi untuk mengontrol ukuran pohon keputusan yang dibuat oleh model *machine learning* agar dapat mencegah terjadinya pembelajaran model *machine learning* yang *overfitting*.

	precision	recall	f1-score	support
dissatisfied	0.93	0.96	0.94	47036
satisfied	0.94	0.90	0.92	36087
accuracy			0.93	83123
macro avg	0.93	0.93	0.93	83123
weighted avg	0.93	0.93	0.93	83123

Gambar 15 Hasil evaluasi performa model *Decision Tree* pada data *training*

Pada Gambar 15 dapat diketahui bahwa hasil evaluasi metrik pada data *training* model *machine learning* *Decision Tree* memiliki nilai presisi *satisfied* sebesar 94% dan *dissatisfied* sebesar 93%. Nilai akurasi pada data *training* model *machine learning* *Decision Tree* juga dinilai sudah cukup bagus karena nilai akurasinya sebesar 93%.

	precision	recall	f1-score	support
dissatisfied	0.91	0.97	0.94	11843
satisfied	0.95	0.87	0.91	8938
accuracy			0.93	20781
macro avg	0.93	0.92	0.92	20781
weighted avg	0.93	0.93	0.93	20781

Gambar 16 Hasil evaluasi performa model *Decision Tree* pada data *testing*

Pada Gambar 16 dapat diketahui bahwa hasil evaluasi metrik pada data *testing* model *machine learning* *Decision Tree* memiliki nilai presisi *satisfied* sebesar 95% dan *dissatisfied* sebesar 91%. Nilai akurasi pada data *testing* model *machine learning* *Decision Tree* juga dinilai memiliki performa yang baik karena nilai akurasinya sebesar 93%. Hal ini menunjukkan bahwa model *machine learning* yang dibuat sudah dapat melakukan pembelajaran dengan baik sehingga hasil performa model yang dihasilkan tidak *underfitting* ataupun *overfitting*.

3. Random Forest

Proses pembuatan model *machine learning* *Random Forest* pada penelitian ini menggunakan nilai *max_depth* sebesar 7, hal ini menunjukkan pada *Random Forest* kedalaman maksimum pada *tree* bernilai 7 maka dilakukanlah Analisa perbandingan performa pada data *training* dan data *testing* seperti yang dapat dilihat pada Gambar 17 dan Gambar 18.

Pada Gambar 17 dapat diketahui bahwa hasil evaluasi metrik pada data *training* model *machine learning* *Random Forest* memiliki nilai presisi *satisfied* sebesar 94% dan *dissatisfied* sebesar 93%. Nilai akurasi pada data *training* model *machine learning* *Random Forest* juga dinilai sudah cukup bagus karena nilai akurasinya sebesar 93%.

	precision	recall	f1-score	support
dissatisfied	0.94	0.95	0.94	47036
satisfied	0.93	0.92	0.92	36087
accuracy			0.93	83123
macro avg	0.93	0.93	0.93	83123
weighted avg	0.93	0.93	0.93	83123

Gambar 17 Hasil evaluasi performa model *Random Forest* pada data *training*

	precision	recall	f1-score	support
dissatisfied	0.93	0.95	0.94	11843
satisfied	0.93	0.91	0.92	8938
accuracy			0.93	20781
macro avg	0.93	0.93	0.93	20781
weighted avg	0.93	0.93	0.93	20781

Gambar 18 Hasil evaluasi performa model *Random Forest* pada data *testing*

Pada Gambar 18 dapat diketahui bahwa hasil evaluasi metrik pada data *testing* model *machine learning* *Random*

Forest memiliki nilai presisi yang sama antara penumpang *satisfied* dan *dissatisfied* yaitu sebesar 93%. Nilai akurasi pada data *testing* model *machine learning Random Forest* juga dinilai memiliki performa yang baik karena nilai akurasinya sebesar 93%. Hal ini menunjukkan bahwa model *machine learning* yang dibuat sudah dapat melakukan pembelajaran dengan baik sehingga hasil performa model yang dihasilkan tidak *underfitting* ataupun *overfitting* karena performanya pada data *training* dan pada data *testing* hamper memiliki kemiripan.

TABEL II
HASIL PERBANDINGAN NILAI MODEL MACHINE LEARNING

Data	KNN	Decision Tree	Random Forest
Training	Accuracy: 94.34%	Accuracy: 93.16%	Accuracy: 93.56%
	Precision: 96.64%	Precision: 94.07%	Precision: 93.25%
	Recall: 90.09%	Recall: 89.9%	Recall: 91.81%
	F1-Score: 93%	F1-Score: 91%	F1-Score: 92%
Testing	Accuracy: 92.61%	Accuracy: 92.96%	Accuracy: 93.22%
	Precision: 96%	Precision: 93.69%	Precision: 92.84%
	Recall: 87.41%	Recall: 89.67%	Recall: 91.27%
	F1-Score: 91%	F1-Score: 91%	F1-Score: 91%

Setelah membuat tiga model *machine learning* dan melakukan evaluasi metrik kepada ketiga model tersebut, dapat diputuskan bahwa model yang paling optimal adalah model *Random Forest* dan *Decision Tree* hal ini dapat dilihat berdasarkan Tabel II. Keputusan tersebut diambil berdasarkan nilai evaluasi metrik *Accuracy* dan *Recall* dari tiap-tiap model merupakan yang tertinggi. Penilaian tersebut didasari pada dataset *Airline Passenger Satisfaction* yang memiliki jumlah data *false negative* dan *false positive* yang sangat mendekati atau simetris di kedua model tersebut. Hal ini perlu dilakukan pembuktian dengan metode validasi.

D. Validation Technique

Untuk memastikan hasil pembuatan model *machine learning* berjalan dengan optimal, Penelitian ini menggunakan *Validation Technique* yang berfokus pada metode *cross validation*. Hal ini dilakukan untuk memastikan bahwa model *machine learning* yang dibuat sudah dapat diimplementasikan untuk menguji data *testing* yang lainnya. Pada penelitian ini, *Validation technique* dilakukan hanya untuk model *machine learning* yang dipilih dan dinilai baik berdasarkan hasil evaluasi metrik yang telah dilakukan.

TABEL III
HASIL CROSS VALIDATION

Parameter	Decision Tree	Random Forest
Accuracy	94.5%	96%
Precision	93.4%	97%
Recall	94%	93.9%
F1-Score	93.7%	95.5%

Cross validation merupakan sebuah metode yang digunakan untuk melakukan evaluasi terhadap kinerja prediktif dari model *machine learning*. Berdasarkan hasil

pada Tabel III, dapat dilihat bahwa hasil *cross validation* untuk model *decision tree* dan *random forest* sudah sangat bagus jika dilihat dari berbagai parameter yang ada. Sehingga bisa dikatakan model *machine learning* yang telah dibuat sudah optimal untuk memprediksi kepuasan pelanggan maskapai penerbangan dan dapat diterapkan sebagai alat bantu untuk mengklasifikasi nilai kepuasan pelanggan maskapai penerbangan x tanpa harus menganalisis kepuasan pelanggan secara manual menggunakan tenaga manusia.

IV. KESIMPULAN

Ada beberapa layanan maskapai penerbangan X yang harus ditingkatkan lagi kualitas pelayanannya, seperti *baggage handling*, *Inflight wifi service*, *departure/arrival time convenient*, *ease of online booking*, dan *online boarding*. Peningkatan kualitas pelayanan perlu dilakukan oleh maskapai penerbangan X agar dapat meningkatkan jumlah keuntungan untuk maskapai penerbangan karena banyaknya pelanggan yang merasa puas dengan pelayanan yang diberikan. Selain itu, model *machine learning* yang telah dibuat pada penelitian ini berhasil berjalan dengan sangat efektif karena memiliki hasil *cross validation* yang bagus. Sehingga model *machine learning Decision Tree* dan *Random Forest* dapat diterapkan oleh maskapai penerbangan X untuk membantu dalam menentukan nilai kepuasan pelanggan tanpa harus menganalisis kepuasan pelanggan secara manual menggunakan tenaga manusia.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan kepada PT. Digitalisasi Pemuda Indonesia karena telah memberikan dataset *Airline Passenger Satisfaction* untuk dilakukan project penelitian *machine learning* ini.

DAFTAR RUJUKAN

- [1] Widjaja, E. L., & Harianto, A. (2017). Analisa Pengaruh Kualitas Layanan Terhadap Kepuasan Penumpang Maskapai Penerbangan Batik Air. *Jurnal Hospitality dan Manajemen Jasa*, 5(2).
- [2] Hayadi, B., Kim, J., Hulliyah, K., & Sukmana, H. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. *International Journal of Informatics and Information Systems*, 4(1), 82-94. doi:<https://doi.org/10.47738/ijjis.v4i1.80>.
- [3] Maryanto, B. (2017). Big Data dan Pemanfaatannya dalam Berbagai Sektor. *Media Informatika*, 16(2), 14-19.
- [4] Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.
- [5] Moro, S., Esmerado, J., Ramos, P., & Alturas, B. (2019). Evaluating a guest satisfaction model through data mining. *International Journal of Contemporary Hospitality Management*.

- [6] Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. *J. Teknol. Indones*, 3.
- [7] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang. *IJCIT (Indonesian J. Comput. Inf. Technol.*, 5(1), 75-82.
- [8] Thupae, R., Isong, B., Gasela, N., & Abu-Mahfouz, A. M. (2018, October). Machine learning techniques for traffic identification and classification in SDWSN: A survey. In *IECON 2018-44th annual conference of the IEEE Industrial Electronics Society* (pp. 4645-4650). IEEE.
- [9] Das, S., & Nene, M. J. (2017, March). A survey on types of machine learning techniques in intrusion prevention systems. In *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 2296-2299). IEEE.
- [10] Krisandi, N., & Helmi, B. P. (2013). Algoritma k-Nearest Neighbor dalam Klasifikasi Data Hasil Produksi Kelapa Sawit pada PT. Minamas Kecamatan Parindu. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 2(1).
- [11] Pulungan, W. A., Mulyani, Y., & Sulistiono, W. E. (2019). Identifikasi Kematangan Buah Kopi Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantization. *Barometer*, 4(2), 217-219.
- [12] Hayadi, B., Kim, J., Hulliyah, K., & Sukmana, H. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. *International Journal of Informatics and Information Systems*, 4(1), 82-94. doi:<https://doi.org/10.47738/ijis.v4i1.80>.
- [13] <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>, diakses Oktober 2021.
- [14] <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>, diakses Januari 2022.
- [15] Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269-274.
- [16] Sutoyo, I. (2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Pilar Nusa Mandiri: Journal of Computing and Information System*, 14(2), 217-224.
- [17] Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(1), 27-31.
- [18] Religia, Y., & Amali, A. (2021). Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 527-533.
- [19] Naufal, M. F. (2021). Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn untuk Klasifikasi Citra Cuaca. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(2), 311-317.