

KLASIFIKASI TINGKAT KEPUASAN DI MASKAPAI PENERBANGAN: STUDI KOMPARASI ALGORITMA K-NN DAN ADABOOST

Ade Maysa¹, Syarifah Putri Agustini Alkadri², Istikoma³

^{1,2} Jurusan Teknik Informatika, ³ Jurusan Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pontianak

¹ademaysaa@gmail.com, ² agustini.putri@unmuhpnk.ac.id, ³ istikoma@unmuhpnk.ac.id

Abstrak

Dalam persaingan bisnis yang ketat, kepuasan pelanggan menjadi kunci utama dalam meningkatkan kinerja perusahaan, terutama di industri penerbangan. Faktor-faktor seperti bagasi tertinggal di bandara keberangkatan, *air conditioning* (AC) pesawat yang tidak berfungsi selama penerbangan, dan keterlambatan penerbangan hingga 2 jam berpotensi besar mempengaruhi kepuasan penumpang. Penelitian ini dilakukan untuk membantu perusahaan melakukan survei guna memahami faktor-faktor yang memengaruhi kepuasan pelanggan dalam mencapai keunggulan kompetitif dan meningkatkan efisiensi layanan. Penelitian ini menggunakan dua metode klasifikasi untuk menganalisis data kepuasan pelanggan, dimana kinerja dari kedua metode tersebut akan dibandingkan. Model klasifikasi yang akan dikomparasi adalah K-Nearest Neighbors (K-NN) dan Adaptive Boosting (Adaboost). Kedua algoritma diuji dengan teknik *k-fold cross validation* dan dievaluasi menggunakan *confusion matrix* untuk mengukur performa masing-masing. Sampel data sebanyak 129.880 diambil dari website Kaggle.com dan dilakukan *preprocessing* sebelum analisis untuk membangun model. Hasil akurasi algoritma K-NN sebesar 86%, dengan nilai *precision* 85% dan *recall* 78%. Sementara itu, Adaboost berhasil mencapai akurasi 90%, dengan nilai *precision* 90% dan *recall* 84%. Dengan akurasi, *precision*, dan *recall* yang lebih tinggi, Adaboost menjadi pilihan terbaik, dengan mencatat nilai *true positive* (TP) 3.199 dan *true negative* (TN) 5.550. Penelitian ini diharapkan dapat membantu perusahaan penerbangan dalam meningkatkan kualitas layanan dan kepuasan pelanggan melalui penggunaan algoritma Adaboost yang terbukti lebih akurat dalam mengklasifikasikan tingkat kepuasan penumpang.

Kata kunci : penerbangan, kepuasan penumpang, K-NN, Adaboost, data mining, klasifikasi

1. Pendahuluan

Dalam era modern ini, mobilitas manusia yang meningkat pesat telah memberikan dampak signifikan pada sektor transportasi di berbagai bidang, termasuk darat, laut, dan udara. Alat transportasi, khususnya transportasi udara, memainkan peran vital dalam menghubungkan berbagai wilayah dengan cepat, sehingga menjadi pilihan utama bagi banyak individu (Aditiya & Latifa, 2023). Dalam persaingan bisnis yang semakin ketat, kepuasan pelanggan menjadi faktor kunci yang menentukan performa perusahaan, terutama dalam industri penerbangan (Yuliarina & Hendry, 2022). Berbagai aspek dapat mempengaruhi tingkat kepuasan penumpang, seperti bagasi yang tertinggal di bandara keberangkatan (Iswara, 2023), sistem *Air Conditioning* (AC) pesawat yang tidak berfungsi saat penerbangan (Ramli, 2023), dan keterlambatan penerbangan selama dua jam (Hikam, 2023). Faktor-faktor ini sangat mempengaruhi pengalaman dan kepuasan penumpang. Perusahaan penerbangan perlu melakukan survei untuk memahami faktor-faktor yang mempengaruhi kepuasan pelanggan guna mencapai keunggulan kompetitif melalui peningkatan efisiensi layanan (Gultom et al., 2018).

Mengukur kualitas layanan dengan indikator kepuasan penumpang adalah penting dalam pandangan bisnis. Dengan menggunakan pendekatan *data mining*, khususnya model klasifikasi, perusahaan dapat mengidentifikasi komponen utama yang berkontribusi terhadap kepuasan atau ketidakpuasan pelanggan, seperti kualitas makanan, ketepatan waktu penerbangan, dan layanan di dalam pesawat (Yoga Religia & Amali, 2021).

Beberapa algoritma yang sering digunakan untuk klasifikasi tingkat kepuasan penumpang adalah K-Nearest Neighbor (K-NN) dan Adaptive Boosting (Adaboost). Algoritma K-NN telah terbukti memberikan akurasi tinggi dalam beberapa penelitian sebelumnya. Misalnya, penelitian "Komparasi Algoritma Decision Tree, Random Forest, SVM, dan K-NN dalam Klasifikasi Kepuasan Penumpang Maskapai Penerbangan" menunjukkan bahwa algoritma K-NN mencapai akurasi sebesar 93% (Setiono, 2022). Penelitian lainnya menunjukkan bahwa algoritma K-NN dengan parameter $K = 7$ dan penggunaan jarak *Euclidean* yang dinormalisasi menghasilkan akurasi sebesar 94% (Aditiya & Latifa, 2023). Penelitian dengan pendekatan optimasi *Backward Elimination* untuk klasifikasi kepuasan

pelanggan menggunakan algoritme K-NN dan Naïve Bayes menunjukkan bahwa K-NN mencapai akurasi sebesar 97.28% dengan nilai *Area Under the Curve* (AUC) sebesar 0.966 (Yunitasari et al., 2021). Di sisi lain, algoritma Adaboost juga menunjukkan performa yang signifikan dalam beberapa studi. Misalnya, penelitian “Perbandingan algoritma untuk analisis sentimen pada twitter transportasi umum *commuterline*” (Novaneliza et al., 2023). Dalam hal ini, algoritma NB-Adaboost mencapai akurasi 78.80%. Studi lainnya membahas “*Machine-Learning Techniques for Customer Retention: A Comparative Study*” menunjukkan bahwa Adaboost mencapai akurasi sebesar 96% (Sabbeh, 2018). Studi lainnya membahas “*Customer Satisfaction Recognition Based on Facial Expression and Machine Learning Techniques*” dalam penelitian ini Adaboost memperoleh akurasi sebesar 98.66% (Bouzakraoui et al., 2020). Namun, belum banyak penelitian yang secara khusus membahas perbandingan antara algoritma K-NN dan Adaboost dalam konteks klasifikasi kepuasan penumpang maskapai penerbangan.

Penelitian ini akan melakukan komparasi antara algoritma K-NN dan Adaboost untuk menentukan algoritma yang memiliki performa terbaik dalam mengklasifikasikan tingkat kepuasan penumpang. Pembaharuan dalam penelitian ini dilakukan dengan menerapkan seleksi fitur dan pengujian setiap algoritma menggunakan teknik *k-fold cross validation* untuk memastikan generalisasi model yang lebih baik. Teknik confusion matrix akan digunakan untuk mengukur performa dari kedua algoritma, yang membantu dalam menentukan seberapa akurat klasifikasi yang dihasilkan oleh masing-masing algoritma (Hendrastuty et al., 2021). Dengan mengetahui algoritma terbaik, perusahaan penerbangan dapat mengimplementasikan metode tersebut untuk meningkatkan kepuasan pelanggan dan mencapai keunggulan kompetitif. Penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai faktor-faktor yang mempengaruhi tingkat kepuasan penumpang maskapai penerbangan dan membantu perusahaan dalam mengklasifikasikan faktor-faktor tersebut dengan lebih akurat dan efisien.

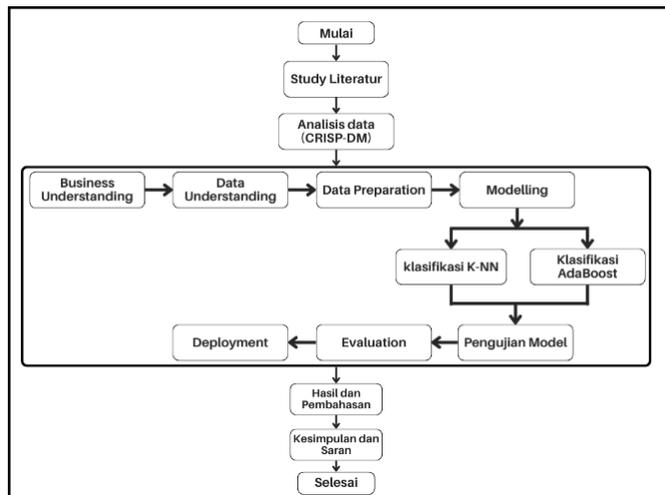
2. Metode Penelitian

2.1 Tahapan Penelitian

Dalam penelitian akan menjabarkan beberapa tahapan. Detail dari tahapan-tahapan penelitian ini dapat di lihat pada gambar 1.

2.2 Business Understanding

Tujuan bisnis dari penelitian ini adalah mengetahui faktor-faktor yang mempengaruhi tingkat kepuasan penumpang di maskapai penerbangan, dengan melakukan klasifikasi tingkat kepuasan penumpang di maskapai penerbangan dapat menjadi acuan perusahaan meningkatkan pelayanan meraka menjadi lebih baik lagi.



Gambar 1. Tahapan Penelitian

2.3 Data Understanding

Pada tahap *data understanding* akan dilakukannya pengumpulan data, dan menganalisa data untuk memahami data yang digunakan. Untuk melaksanakan penelitian ini, dataset *Airlines Passenger Satisfaction* yang bersumber dari *kaggle.com* tahun 2022 menjadi pilihan penulis, dataset tersebut berjumlah sebanyak 129.880 baris (Ahmad Bhat Mysar, 2022).

2.4 Data Preparation

1. Preprocessing

Hal yang harus dilakukan adalah membersihkan atau *preprocessing* data dari berbagai macam noise yang telah ditemukan pada saat proses analisis dataset yang digunakan. Hal ini dilakukan supaya dataset yang akan digunakan memiliki dataset yang baik, beragam dan layak untuk digunakan. Proses membersihkan dataset sangat berpengaruh terhadap hasil dalam proses klasifikasi. Dalam fase ini ada beberapa proses yang akan dilakukan, seperti dituliskan pada bagian di bawah ini.

a) Seleksi fitur

Atribut atau fitur yang memiliki korelasi rendah dengan label akan di eliminasi agar dapat meningkatkan performa dari model yang akan dibangun nantinya. Untuk melakukan seleksi fitur di gunakan teknik korelasi pearson. Korelasi Pearson adalah teknik statistik yang digunakan untuk mengukur kekuatan dan arah hubungan linear antara dua variabel. Ini menentukan seberapa dekat dua variabel dengan satu sama lain. Koefisien korelasi Pearson berkisar dari nol hingga satu. Nilai nol menunjukkan korelasi sempurna negatif, nol menunjukkan tidak ada korelasi, dan satu menunjukkan korelasi sempurna positif. Korelasi -0,1 menunjukkan hubungan negatif yang lemah dan korelasi -0,8 menunjukkan hubungan negatif yang kuat (Nikmatun & Waspada, 2019).

b) Identify Duplicate Rows

Tahap ini dilakukan untuk mengetahui apakah terdapat baris data yang duplikat. Jika terdapat baris data yang duplikat maka akan di hapus agar dataset memiliki data yang banyak dan beragam.

c) Identiy Missing or Null Values

Selanjutnya akan dilakukan pengecekan untuk mengetahui apakah terdapat baris data yang mempunyai nilai kosong atau *null*.

d) Outliers Detection

Berikutnya, akan dilakukan pengecekan untuk menentukan apakah fitur memiliki nilai yang berbeda secara umum. Hal ini dilakukan karena nilai *outlier* pada fitur akan mengurangi nilai perhitungan mean karena rata-rata yang dihasilkan tidak dapat menggambarkan data secara keseluruhan.

e) Data Imbalance

Jika terdapat ketidak seimbangan dalam data label maka peneliti akan mengatasinya dengan teknik *random over sampler*.

f) Standarisasi

Jika pada tahap sebelumnya ditemukan nilai *outliers* pada data. Maka harus di tangani dengan melakukan normalisasi data dengan teknik standarisasi data. Penulis akan menggunakan *library* Standar Scaler untuk melakukan standarisasi data.

2.5 Modelling

Pada tahap *modelling* yang akan dilakukan proses klasifikasi data menggunakan algoritma K-Nearest Neighbors (K-NN) dan algoritma Adaptive Boosting (Adaboost).

2.6 Pengujian Model

Proses selanjutnya yaitu pengujian model dimana kedua model akan di uji dengan teknik K-fold Cross Validation dengan nilai k=10 (Koenker et al., 2017).

2.7 Evaluation

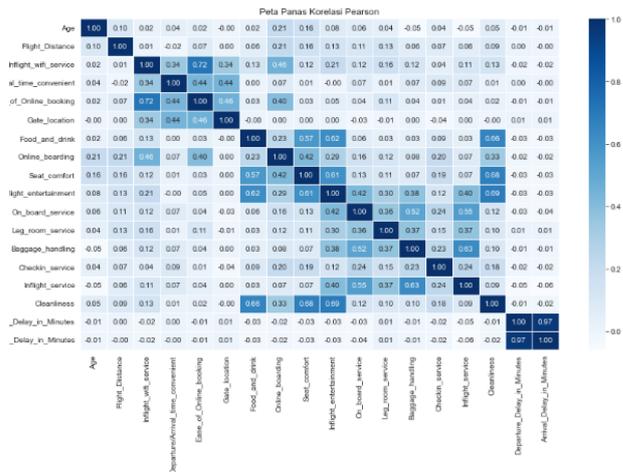
Tahap ini adalah tahap untuk mengetahui kinerja algoritma dengan menghitung *confusion matrix* untuk melihat nilai *accuracy, recall, precision*.

3. Hasil dan Pembahasan

3.1 Hasil Preprocessing

1. Seleksi Fitur

Berdasarkan heatmap korelasi pearson pada gambar 2, didapatkan fitur-fitur yang akan penulis gunakan untuk proses pemodelan. Sementara untuk kolom ID, *age, gender, flight distance, gate location, online boarding, leg room service, check in service, departure and arrival time convenience* akan di *drop*. Adapun fitur yang digunakan pada penelitian ini dapat dilihat pada tabel 1.



Gambar 2. Heatmap Korelasi Pearson

Tabel 1. Seleksi Fitur

No.	Nama fitur	Keterangan
1.	Customer Type	Jenis pelanggan First Time (0), Returning (1)
2.	Type of Travel	Tujuan penerbangan penumpang Perjalanan pribadi(0), Perjalanan Bisnis(1)
3.	Class	Kelas perjalanan penumpang Bisnis (0), Eco (1), Eco Plus (2)
4.	Inflight Wifi Service	Tingkat kepuasan layanan wifi dalam pesawat(1-5)
5.	Ease of Online Booking	Tingkat kepuasan pemesanan Online (1-5)
6.	Food and Drink	Tingkat kepuasan makanan dan minuman (1-5)
7.	Seat Comfort	Tingkat kepuasan kenyamanan kursi (1-5)
8.	Inflight Entertainment	Tingkat kepuasan hiburan dalam pesawat (1-5)
9.	On-Board Service	Tingkat kepuasan layanan On-board (1-5)
10.	Baggage Handling	Tingkat kepuasan penanganan bagasi (1-5)
11.	Inflight Service	Tingkat kepuasan layanan dalam pesawat (1-5)
12.	Cleanliness	Tingkat kepuasan kebersihan (1-5)
13.	Departure Delay in Minutes	Menit ditunda saat keberangkatan
14.	Arrival Delay in Minutes	Menit ditunda saat kedatangan

2. Label Encode

Peneliti melakukan pengecekan tipe data dari setiap fitur dan label yang digunakan dalam penelitian, tipe data dari fitur dan label dapat dilihat pada gambar 3.

Berdasarkan gambar 3, terdapat tipe data yang harus di label *encodes* sehingga dapat diproses oleh *machine learning*, dalam penelitian ini penulis melakukan label encode terhadap 4 variabel bertipe data *object*.

Dan gambar 4 adalah data yang telah dilakukan label encode dari kategorik menjadi numerik sehingga dapat diproses ke tahap selanjutnya.

Data columns (total 15 columns):			
#	Column	Non-Null Count	Dtype
0	Customer Type	129487 non-null	object
1	Type of Travel	129487 non-null	object
2	Class	129487 non-null	object
3	Departure Delay	129487 non-null	int64
4	Arrival Delay	129487 non-null	float64
5	Ease of Online Booking	129487 non-null	int64
6	On board Service	129487 non-null	int64
7	Seat Comfort	129487 non-null	int64
8	Cleanliness	129487 non-null	int64
9	Food and Drink	129487 non-null	int64
10	In flight Service	129487 non-null	int64
11	In flight Wifi Service	129487 non-null	int64
12	In flight Entertainment	129487 non-null	int64
13	Baggage Handling	129487 non-null	int64
14	Satisfaction	129487 non-null	object

Gambar 3. Tipe Data Fitur dan Label

Data columns (total 15 columns):			
#	Column	Non-Null Count	Dtype
0	Customer Type	129487 non-null	int32
1	Type of Travel	129487 non-null	int32
2	Class	129487 non-null	int32
3	Departure Delay	129487 non-null	float64
4	Arrival Delay	129487 non-null	float64
5	Ease of Online Booking	129487 non-null	int64
6	On board Service	129487 non-null	int64
7	Seat Comfort	129487 non-null	int64
8	Cleanliness	129487 non-null	int64
9	Food and Drink	129487 non-null	int64
10	In flight Service	129487 non-null	int64
11	In flight Wifi Service	129487 non-null	int64
12	In flight Entertainment	129487 non-null	int64
13	Baggage Handling	129487 non-null	int64
14	Satisfaction	129487 non-null	int32

Gambar 4. Tipe Data setelah Encode.

3. Missing Value

Setelah melakukan pengecekan untuk memastikan apakah dataset memiliki nilai yang hilang, hasilnya menunjukkan bahwa di kolom *arrival delay* memiliki nilai kosong, seperti yang ditunjukkan pada gambar 5.

	Total data yang hilang	Persentase data yang hilang
Departure Delay	0	0.000
Ease of Online Booking	0	0.000
On board Service	0	0.000
Seat Comfort	0	0.000
Cleanliness	0	0.000
Food and Drink	0	0.000
In flight Service	0	0.000
In flight Wifi Service	0	0.000
In flight Entertainment	0	0.000
Baggage Handling	0	0.000
Satisfaction	0	0.000

Gambar 5. Missing Value

Tabel diatas menunjukkan bahwa terdapat 393 data yang hilang pada kolom *arrival delay* yang mana harus ditangani, pada penelitian ini penulis menangani data yang hilang dengan cara mengisi data tersebut dengan nilai median dari data set agar tidak

mempengaruhi proses berikutnya. Untuk hasilnya bisa dilihat pada gambar 6.

	Total data yang hilang	Persentase data yang hilang
Arrival Delay	0	0.0
Customer Type	0	0.0
Type of Travel	0	0.0
Class	0	0.0
Departure Delay	0	0.0
Ease of Online Booking	0	0.0
On board Service	0	0.0
Seat Comfort	0	0.0
Cleanliness	0	0.0
Food and Drink	0	0.0
In flight Service	0	0.0
In flight Wifi Service	0	0.0
In flight Entertainment	0	0.0
Baggage Handling	0	0.0
Satisfaction	0	0.0

Gambar 6. Menangani Missing Value

4. Data Duplicates

Setelah melakukan pengecekan dengan *python*, penulis menemukan bahwa terdapat 33,068 baris data yang terdapat duplikasi yang dapat dilihat pada gambar 7.

Customer Type	Type of Travel	Class	Departure Delay	Arrival Delay	Ease of Online Booking	On board Service	Seat Comfort	Cleanliness
171	Returning	Business	Business	0	0.000000	3	5	5
336	Returning	Personal	Economy	0	0.000000	2	4	5
356	Returning	Business	Business	0	0.000000	2	5	5
358	Returning	Business	Business	0	0.000000	0	5	4
360	Returning	Business	Business	0	0.000000	5	5	5
129863	Returning	Business	Business	0	0.000000	5	5	4
129864	Returning	Business	Business	0	0.000000	3	5	4
129868	First-time	Business	Economy	0	0.000000	2	4	5
129873	Returning	Personal	Economy Plus	0	0.000000	4	1	3
129876	Returning	Personal	Economy Plus	0	0.000000	3	5	2

Number of rows: 33068
Number of columns: 15

Gambar 7. Data Duplikat pada Dataset

Untuk menjaga agar dataset memiliki data yang banyak dan beragam, penulis menghapus 33,068 baris data tersebut dengan cara *drop duplicate*.

5. Outliers

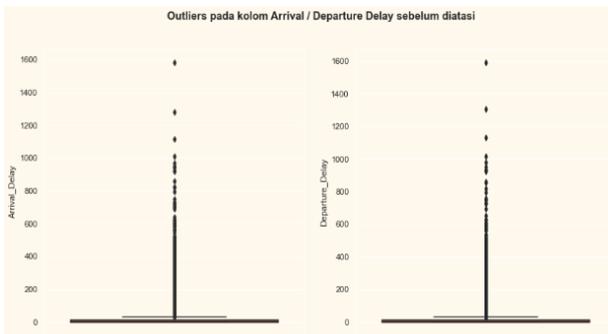
Berdasarkan hasil dari pengecekan terdapat 4 fitur yang memiliki nilai outliers, yaitu *arrival delay*, *departure delay*, *baggage handling*, *inflight service* seperti yang dapat dilihat pada gambar 9.

Column Name	% Outliers	
0	Arrival Delay	11.1050
1	Baggage Handling	8.2417
2	Class	0.0000
3	Cleanliness	0.0000
4	Customer Type	0.0000
5	Departure Delay	10.9480
6	Ease of Online Booking	0.0000
7	Food and Drink	0.0000
8	In flight Entertainment	0.0000
9	In flight Service	8.1302
10	In flight Wifi Service	0.0000
11	On board Service	0.0000
12	Satisfaction	0.0000
13	Seat Comfort	0.0000
14	Type of Travel	0.0000

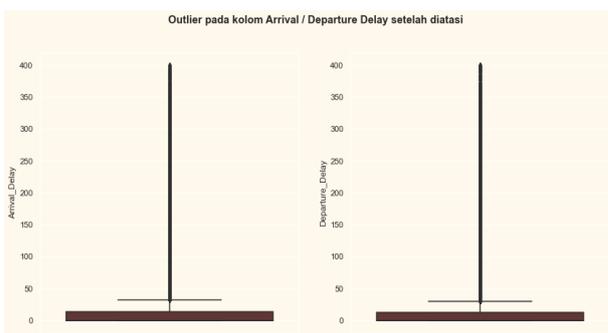
Gambar 9. Deteksi outliers pada dataset

Outliers pada ke 4 kolom tersebut harus diatasi, sehingga penulis melakukan penanganan pada ke 4

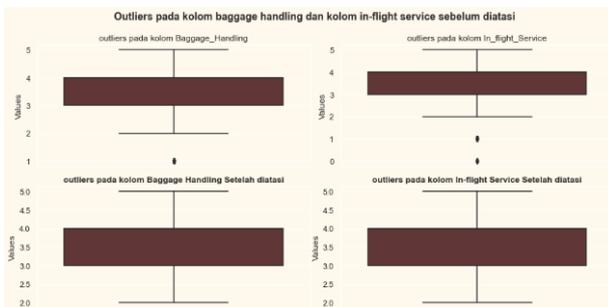
kolom tersebut dengan merubah data *outliers* menjadi nilai *mean* dari data.



Gambar 10. Outliers pada Arrival/Departure Delay



Gambar 11. Menangani Outliers



Gambar 12. Menangani Outliers pada Baggage Handling dan Inflight Service

Gambar 10, 11 dan 12 adalah proses penanganan nilai outliers pada setiap kolom sehingga dapat dilakukan proses selanjutnya.

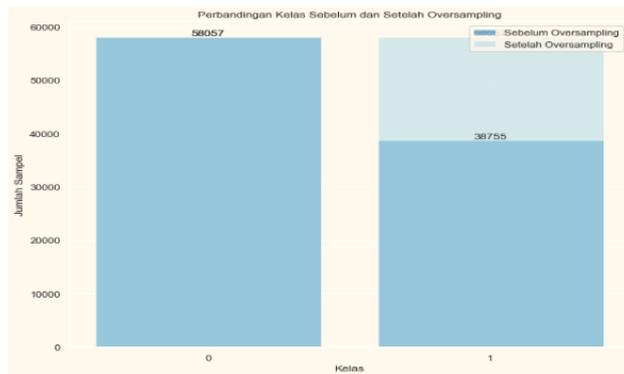
6. Data Imbalance

Setelah dilakukan pengecekan data imbalance pada dataset diketahui bahwa data label 0 (tidak puas/netral) dan data label 1 (puas) ternyata kelasnya tidak seimbang yang dapat dilihat pada gambar 13.

Sebelum dilakukan *over sampling* dengan teknik *random over sampler* data 0 (tidak puas/netral) berjumlah 58.057 dan data 1 (Puas) berjumlah 38.755. Dan setelah dilakukan *over sampling* data 0 dan 1 sama-sama berjumlah 58.057.

7. Standarisasi

Pada penelitian ini penulis melakukan proses standarisasi menggunakan *standar scaler*. Hasil dari standarisasi dapat dilihat pada gambar 14.



Gambar 13. Barplot Kelas Sebelum dan Sesudah Oversampling

	mean	min	max	median	std
Customer_Type	0.000000	2.039070	0.490420	0.490420	1.000005
Type_of_Travel	0.000000	0.693972	1.440980	0.693972	1.000005
Class	0.000000	1.020568	2.118664	0.549058	1.000005
Departure_Delay	0.000000	0.494535	9.852781	0.416930	1.000005
Arrival_Delay	0.000000	0.502446	9.784730	0.399574	1.000005
Ease_of_Online_Booking	0.000000	1.928260	1.601585	0.189647	1.000005
On_board_Service	0.000000	2.517010	1.308047	0.221976	1.000005
Seat_Comfort	0.000000	2.533474	1.230633	0.477811	1.000005
Cleanliness	0.000000	2.440768	1.332115	0.177038	1.000005
Food_and_Drink	0.000000	2.371675	1.357872	0.133947	1.000005
In_flight_Service	0.000000	1.838565	1.320067	0.267189	1.000005
In_flight_Wifi_Service	0.000000	2.033669	1.716037	0.216155	1.000005
In_flight_Entertainment	0.000000	2.458177	1.271594	0.220314	1.000005
Baggage_Handling	0.000000	1.830343	1.327045	0.274583	1.000005
Satisfaction	0.000000	0.817027	1.223949	0.817027	1.000005

Gambar 14. Standarisasi Data

Gambar 14 adalah hasil dari standarisasi menggunakan standar *scaler*, tahapan ini dilakukan sebelum membangun model.

3.2 Hasil Pengujian Model

1. Pengujian Algoritma K-NN

Pada pengujian algoritma K-NN melakukan *preprocessing* pada dataset yang terdiri dari seleksi fitur, mengisi *missing value*, mengecek data duplikat, melakukan label encoding dan mengatasi *outliers*, menangani *data imbalance*, melakukan standarisasi. Pada penelitian ini menggunakan nilai *k-fold* = 10 sehingga terdapat 10 iterasi yang dapat dilihat pada tabel 2.

Tabel 2. Akurasi Algoritma K-NN per Iterasi

Iterasi ke-	Akurasi
1	0.8541623631481099
2	0.858293740962611
3	0.858072513170127
4	0.8506352649519677
5	0.8584856936266915
6	0.8561099060014461
7	0.8606548910236546
8	0.854973659745894
9	0.8542505939469063
10	0.8617911372792066

2. Pengujian Algoritma Adaboost

Selanjutnya pengujian pada algoritma Adaboost dengan proses preprocessing yang sama dan menggunakan nilai *k-fold* = 10 diperoleh akurasi seperti pada tabel 3.

Tabel 3. Akurasi Algoritma Adaboost per Iterasi

Iterasi ke-	Akurasi
1	0.9046684569303863
2	0.9009502168973352
3	0.9065179217023035
4	0.9037289536204938
5	0.9088937093275488
6	0.9021795269083772
7	0.9022828220225183
8	0.9026960024790828
9	0.9051750852184691
10	0.9037289536204938

3. Perbandingan Akurasi Kedua Algoritma

Untuk melihat hasil akurasi dengan *k-fold cross validation* dari kedua algoritma dapat dilihat pada tabel 4.

Tabel 4. Perbandingan Akurasi

Iterasi ke-	K-NN	Adaboost
1	85.41%	90.46%
2	85.82%	90.09%
3	85.80%	90.65%
4	85.06%	90.37%
5	85.84%	90.88%
6	85.61%	90.21%
7	85.06%	90.22%
8	85.49%	90.26%
9	85.42%	90.51%
10	85.17%	90.37%
Rata - rata	85.67%	90.37%

3.3 Hasil Evaluasi Model

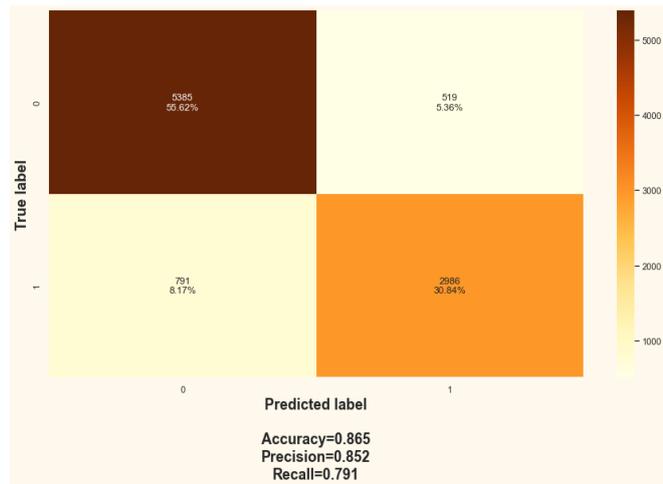
1. Evaluasi Model K-NN

Setelah dilakukan evaluasi model menggunakan *confusion matrix* algoritma K-NN memperoleh *score true positive* (TP) 2.945, *true negative* (TN) 5.398, *false positive* (FP) 506 dan *false negative* (FN) 832. Baik dalam memprediksi kelas 0 dan tidak baik dalam memprediksi kelas 1 yang dapat dilihat pada gambar 15. Perhitungan akurasi K-NN dilakukan dengan menggunakan persamaan (1).

$$\begin{aligned}
 &= \frac{TP + TN}{TP + FP + FN + TN} \quad (1) \\
 &= \frac{2.945 + 5.398}{2.945 + 506 + 832 + 5.398} \\
 &= \frac{8.343}{8.343} \\
 &= \frac{9.681}{9.681} \\
 &= 0.861791137 \times 100 \\
 &= 86\%
 \end{aligned}$$

Pada pengujian ini juga dilakukan perhitungan *Precision* dengan menggunakan persamaan (2).

$$\begin{aligned}
 &= \frac{TP}{TP + FP} \quad (2) \\
 &= \frac{2.945}{2.945 + 506} \\
 &= \frac{3.451}{3.451} \\
 &= 0.853375833 \times 100 = 85\%
 \end{aligned}$$



Gambar 15. Confusion Matrix K-NN

Sedangkan nilai pengujian *recall* dihitung menggunakan persamaan (3).

$$\begin{aligned}
 &= \frac{TP}{TP + FN} \quad (3) \\
 &= \frac{2.945}{2.945 + 832} \\
 &= \frac{3.777}{3.777} \\
 &= 0.779719354 \times 100 \\
 &= 78\%
 \end{aligned}$$

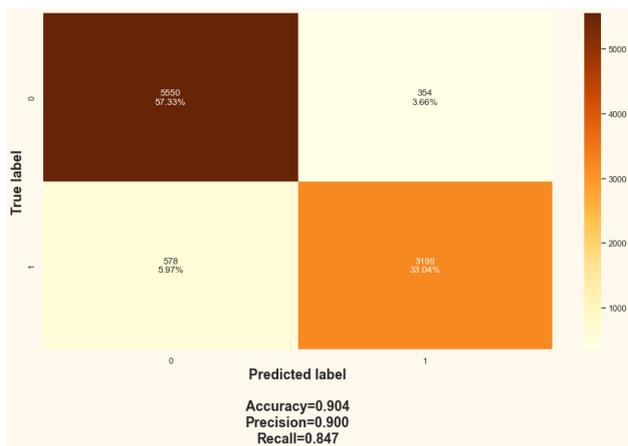
2. Evaluasi Model Adaboost

Setelah dilakukan evaluasi model menggunakan *confusion matrix* algoritma Adaboost memperoleh *score true positive* (TP) 3.199 dan *true negative* (TN) 5.550, *false positive* (FP) 354 dan *false negative* (FN) 578. Baik dalam memprediksi kelas 0 dan tidak baik memprediksi kelas 1 yang dapat dilihat pada Gambar 16. Pengujian pada model Adaboost juga dilakukan dengan menggunakan akurasi yang ditung menggunakan persamaan (1) dan didapatkan nilai akurasi sebesar 90%.

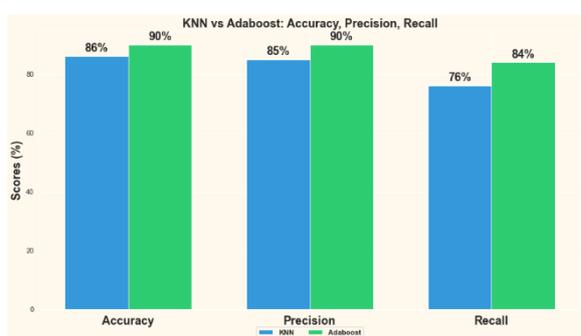
Selain menghitung akurasi, juga dilakukan perhitungan nilai presisi dan *recall*. Perhitungan nilai presisi dilakukan dengan menggunakan persamaan (2), sedangkan perhitungan *recall* dilakukan dengan menggunakan persamaan (3). Nilai presisi yang didapatkan dari perhitungan adalah sebesar 90%, sedangkan nilai *recall* sebesar 84%.

Berdasarkan hasil perhitungan nilai *accuracy*, *precision* dan *recall* yang telah dilakukan sebelumnya, diperoleh perbandingan antara algoritma K-NN dan Adaboost yang dapat dilihat pada gambar 17.

Berdasarkan gambar 17 diperoleh nilai akurasi algoritma K-NN 86%, nilai *precision* 85% dan nilai *recall* 78%. Sedangkan untuk algoritma Adaboost diperoleh nilai akurasi 90%, nilai *precision* 90% dan nilai *recall* 84%. Dimana selisih akurasi antara kedua algoritma sebesar 4%.



Gambar 16. Confusion Matrix Adaboost



Gambar 17. Nilai Accuracy, Precision, Recall K-NN dan Adaboost

4. Kesimpulan dan Saran

Hasil dari penelitian yang dilakukan melalui beberapa tahapan, mulai dari menyeleksi fitur, menangani *missing value*, menangani *outliers*, menangani data *duplicate* melakukan *label encoding*, kemudian menangani data imbalance dengan teknik *random over sampler* dan melakukan standarisasi menggunakan *standar scaler*. dan membangun model klasifikasi dengan algoritma K-NN dan algoritma Adaboost. Hasil dari pengujian akurasi dengan *k-fold cross validation* pada algoritma K-NN dan Adaboost dalam melakukan prediksi terhadap setiap data pada tiap iterasi dan dievaluasi dengan *confusion matrix*. Diperoleh hasil terbaik yaitu algoritma Adaboost dengan nilai akurasi 90%, *precision* 90% dan nilai *recall* 84% dan mendapatkan nilai *true positive* (TP) 3.199 dan nilai *true negative* (TN) 5.550.

Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan mengoptimalkan parameter pada setiap algoritma agar dapat meningkatkan kinerja model, menggunakan teknik evaluasi model lainnya seperti Area Under Curve (AUC) dan menggunakan teknik *oversampling* SMOTE dalam menangani *data imbalance*.

Daftar Pustaka:

Aditiya, D., & Latifa, U. (2023). Uji Efektivitas Penerapan Machine Learning Classification Untuk Survey Kepuasan Pelanggan Maskapai

Penerbangan X. *Barometer*, 8(1), 9–18. <https://doi.org/10.35261/barometer.v8i1.6566>

AHMAD BHAT MYSAR. (2022). *Airline Passenger Satisfaction*. Kaggle.Com. <https://www.kaggle.com/datasets/mysarahmad/airline-passenger-satisfaction>

Bouzakraoui, M. S., Sadiq, A., & Alaoui, A. Y. (2020). Customer satisfaction recognition based on facial expression and machine learning techniques. *Advances in Science, Technology and Engineering Systems*, 5(4), 594–599. <https://doi.org/10.25046/AJ050470>

Gultom, S., Marlita, D., & Wardana, A. (2018). Kualitas Layanan, Kepercayaan Merek dan Kepuasan Penumpang. *Jurnal Manajemen Transportasi & Logistik (JMTRANSLOG)*, 5(2), 169. <https://doi.org/10.54324/j.mtl.v5i2.248>

Hendrastuty, N., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., Isnain, A. R., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <http://situs.com>

Hikam, H. A. Al. (2023). *Fadli Zon Ngeluh Penerbangan Batik Air Yogyakarta-Jakarta Telat 2 Jam Lebih*. Detik.Com. <https://finance.detik.com/berita-ekonomi-bisnis/d-6899447/fadli-zon-ngeluh-penerbangan-batik-air-yogyakarta-jakarta-telat-2-jam-lebih>

Iswara, A. J. (2023). *Sudah Mendarat, Semua Barang Bagasi Pesawat Ini Ketinggalan di Bandara*. Kompas.Com. <https://www.kompas.com/global/read/2023/09/10/234700170/sudah-mendarat-semua-barang-bagasi-pesawat-ini-ketinggalan-di-bandara>

Koenker, R., Chernozhukov, V., He, X., & Peng, L. (2017). Handbook of quantile regression. In *Handbook of Quantile Regression*. <https://doi.org/10.1201/9781315120256>

Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.

Novaneliza, R., Handayani, F., Suhandar, R. J., Suroho, H., Azzahra, N. S., & Nadilla, D. (2023). Perbandingan Algoritma Untuk Analisis Sentimen Pada Twitter Transportasi Umum Commuterline. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(1), 13–21.

Ramli, R. R. (2023). *Tengah Jadi Sorotan gara-gara AC Pesawat Mati Selama Terbang, Siapa Pemilik Super Air Jet?* Kompas.Com.

- <https://money.kompas.com/read/2023/03/24/124200826/tengah-jadi-sorotan-gara-gara-ac-pesawat-mati-selama-terbang-siapa-pemilik>
- Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of Advanced Computer Science and Applications*, 9(2), 273–281. <https://doi.org/10.14569/IJACSA.2018.090238>
- Setiono, M. H. (2022). a Komparasi Algoritma Decision Tree, Random Forest, Svm Dan K-Nn Dalam Klasifikasi Kepuasan Penumpang Maskapai Penerbangan. *INTI Nusa Mandiri*, 17(1), 32–39. <https://doi.org/10.33480/inti.v17i1.3420>
- Yoga Religia, & 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. <https://doi.org/10.29207/resti.v5i3.3086>
- Yuliarina, A. N., & Hendry, H. (2022). Comparison of Prediction Analysis of Gofood Service Users Using the Knn & Naive Bayes Algorithm With Rapidminer Software. *Jurnal Teknik Informatika (Jutif)*, 3(4), 847–856. <https://doi.org/10.20884/1.jutif.2022.3.4.294>
- Yunitasari, Hopipah, H. S., & Mayasari, R. (2021). Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes. *Technomedia Journal*, 6(1), 99–110. <https://doi.org/10.33050/tmj.v6i1.1531>