

# Implementasi Algoritma K-Nearest Neighbors dalam Mengetahui Kepuasan Pengguna Aplikasi CapCut

Alda Maretina Iranti<sup>1</sup>, I Kadek Dwi Nuryana<sup>2</sup>

<sup>1,2</sup> Sistem Informasi, Teknik Informatika, Universitas Negeri Surabaya

<sup>1</sup>[alda.19016@mhs.unesa.ac.id](mailto:alda.19016@mhs.unesa.ac.id)

<sup>2</sup>[dwinuryana@unesa.ac.id](mailto:dwinuryana@unesa.ac.id)

**Abstrak**—Dalam era digital yang terus berkembang, pembuatan konten telah menjadi bidang yang sangat kompetitif dalam penyampaian informasi. Banyak platform dalam bentuk aplikasi untuk membuat suatu konten yang digunakan oleh kebanyakan masyarakat, salah satu aplikasi yang semakin populer adalah CapCut. CapCut menjadi aplikasi kedua yang paling banyak diunduh dan menjadi aplikasi pertama editing video yang paling banyak diunduh di Indonesia pada 2023. CapCut adalah aplikasi pengeditan video yang dirancang untuk memudahkan pengguna dalam membuat konten video yang menarik dan profesional langsung dari perangkat seluler. Dengan munculnya platform ini, muncul pula kebutuhan untuk memahami sejauh mana pengguna puas dengan pengalaman pengeditan video yang mereka dapatkan dari aplikasi ini. Untuk mengetahui kepuasan pengguna dapat menggunakan model teori Unified Theory of Acceptance and Use of Technology (UTAUT). Sedangkan untuk mengidentifikasi faktor yang mempengaruhi kepuasan pengguna digunakan teknik Principal Component Analysis (PCA). Kemudian untuk mengetahui tingkat kepuasan pelanggan menggunakan metode klasifikasi dengan algoritma k-Nearest Neighbor (KNN). Data yang digunakan dalam penelitian ini adalah data kuesioner. Penelitian ini dilakukan dengan beberapa tahap, dimulai dari pengidentifikasian masalah, studi literatur, pengumpulan data, data preprocessing, modeling dan evaluasi. dari penelitian ini didapatkan faktor yang paling mempengaruhi kepuasan pengguna layanan aplikasi CapCut yaitu Effort Expectancy dengan fokus yaitu “Proses layanan yang tidak rumit”, dimana 24 responden atau 11,01% responden merasa TIDAK PUAS dan 194 responden atau 88,99% responden merasa PUAS. Dengan menggunakan metode K-Nearest Neighbors (KNN) dalam penelitian ini mendapatkan hasil akurasi sebesar 0,9878 atau 98,78%. Dan memiliki presisi sebesar 0,9762 atau 97,62%, dan recall 1,00 atau 100%.

**Kata Kunci**— k-NN, PCA, UTAUT, Kepuasan, dan CapCut.

## I. PENDAHULUAN

Dalam era digital yang terus berkembang, pembuatan konten telah menjadi bidang yang sangat kompetitif dalam penyampaian informasi. Meskipun masa depan tidak mungkin diprediksi, pakar industri telah membuat prediksi menarik tentang arah pembuatan konten pada tahun 2024, yang mencakup meningkatnya konten interaktif serta dampak signifikan kecerdasan buatan (AI) dalam menyederhanakan proses pembuatan konten, termasuk pembuatan konten otomatis, rekomendasi personalisasi, dan penyesuaian konten untuk pencarian suara serta asisten cerdas [1]. Banyak platform dalam bentuk aplikasi untuk membuat suatu konten yang digunakan oleh kebanyakan masyarakat, salah satu aplikasi yang semakin populer adalah CapCut.

CapCut adalah aplikasi edit video yang memudahkan pembuatan konten video menarik dan profesional langsung dari perangkat seluler [3]. Dilansir pada laman web databoks.com aplikasi paling banyak diunduh di Indonesia tahun 2023 disajikan pada Gbr. 1 [2].



Gbr. 1 Aplikasi yang Paling Banyak Diunduh di Indonesia (2023)

Seperti yang disajikan pada Gbr. 1, CapCut merupakan aplikasi kedua yang banyak diunduh dan menjadi aplikasi pertama editing video yang paling banyak diunduh di Indonesia pada 2023, dengan total 53,9 juta kali unduhan. CapCut dirancang dengan fitur sederhana yang memudahkan pemula untuk membuat konten video berkualitas tinggi. Popularitas ini mencerminkan tingkat kepuasan pengguna yang tinggi, di mana banyak pengguna menghargai kemudahan penggunaan dan fitur-fitur lengkap yang ditawarkan CapCut. Dengan munculnya platform ini, muncul pula kebutuhan untuk memahami sejauh mana pengguna puas dengan pengalaman pengeditan video yang mereka dapatkan dari aplikasi ini.

Kepuasan pengguna adalah hasil dari bagaimana konsumen menilai perbandingan antara harapan mereka sebelum menggunakan suatu produk atau layanan dengan kinerja aktual produk tersebut setelah digunakan [4]. Keberadaan aplikasi CapCut tentunya harus mampu menjawab kebutuhan pelanggan, karena aplikasi yang usability bisa menjadi pengaruh besar pada kepuasan pelanggan [5]. Oleh sebab itu perlu dilakukannya penelitian dalam mengetahui kepuasan pelanggan pada aplikasi CapCut. Untuk mengetahui kepuasan pengguna digunakan model teori *Unified Theory of Acceptance and Use of Technology* (UTAUT). Penggunaan model ini terbukti mampu menyelesaikan permasalahan dalam mengetahui kepuasan pengguna, seperti yang disampaikan oleh

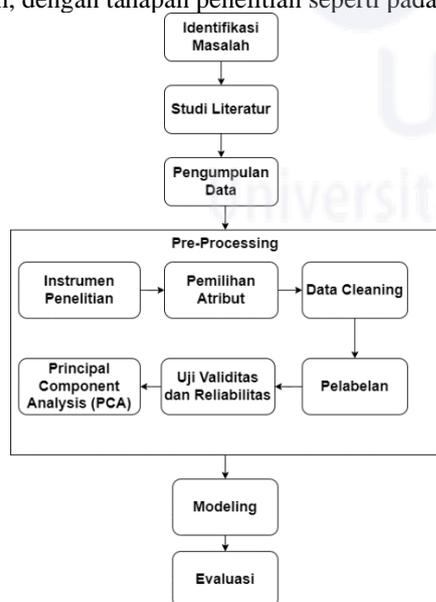
penelitian [6]. Kemudian untuk mengetahui tingkat kepuasan pelanggan digunakan metode klasifikasi dengan algoritma k-Nearest Neighbor (KNN).

K-Nearest Neighbor (KNN) ialah metode regresi non-parametrik yang paling populer dimana dalam metode ini, fungsi distribusi dari nilai prediksi diperoleh dengan menggunakan distribusi non-parametrik dari fungsi kernel [7]. Penggunaan algoritma ini efektif dalam mengetahui tingkat kepuasan pengguna [8]. Sedangkan untuk mengetahui faktor yang mempengaruhi kepuasan pengguna adalah menggunakan teknik *Principal Component Analysis* (PCA). Metode teknik PCA ini mampu digunakan untuk mengembangkan skala pengukuran dalam kuesioner. Penelitian yang menggunakan teknik *Principal Component Analysis* (PCA) yaitu berjudul Analisis Faktor Kepuasan Pelanggan terhadap Layanan Perusahaan Daerah Air Minum [4].

Dari uraian di atas, dalam penelitian ini akan dilakukan pengukuran terhadap faktor yang paling mempengaruhi kepuasan pengguna aplikasi CapCut dengan menggunakan algoritma K-Nearest Neighbors (KNN) untuk mengetahui bagaimana kepuasan pengguna pada layanan aplikasi CapCut. Dengan adanya penelitian ini dapat membantu perusahaan aplikasi CapCut untuk meningkatkan kualitas layanan pada pengguna. Selain itu diharapkan penelitian ini dapat menjadi referensi pada penelitian selanjutnya.

## II. METODE PENELITIAN

Pada penelitian ini menggunakan tools Rapidminer dalam prosesnya. Rapidminer memiliki berbagai operator yang efektif untuk menyelesaikan masalah data mining, terutama dalam klasifikasi yang mendasari metode *machine learning* [9]. Metode yang digunakan dalam penelitian ini yaitu K-Nearest Neighbor (K-NN) dengan teknik *Principal Component Analysis* (PCA) dalam mencari faktor yang paling berpengaruh, dengan tahapan penelitian seperti pada Gbr. 2.



Gbr. 2 Metode Penelitian

Seperti yang dapat dilihat pada Gbr. 2 ada beberapa tahapan penelitian yang akan dilakukan dengan penjelasan secara rinci seperti berikut ini.

### A. Identifikasi Masalah

Pada tahap identifikasi masalah dilakukan perumusan masalah dari obyek penelitian yang telah ditentukan dalam penelitian ini. Rumusan masalah ini akan menjadi tujuan penelitian yang dilakukan.

### B. Studi Literatur

Setelah masalah sudah teridentifikasi, kemudian dilakukan studi literatur untuk mendapatkan dasar penelitian yang kuat. Studi literatur ini mencari referensi-referensi yang terkait pada penelitian ini dari jurnal atau pun dari literatur lain.

### C. Pengumpulan Data

Pada tahap ini, Pengumpulan data akan dilakukan dengan menyebarkan kuesioner menggunakan teknik survei.

Peneliti kuantitatif umumnya mengandalkan dua teknik pengumpulan data, yaitu dengan teknik kuesioner dan tes. Kuesioner lebih banyak digunakan dalam survei, sedangkan tes lebih sesuai untuk penelitian eksperimental. [10]. Kriteria responden yang dibutuhkan pada penelitian ini ialah sebagai berikut :

1. Responden berusia minimal 17 tahun atau lebih
2. Pernah menggunakan aplikasi CapCut minimal 1 kali

Dalam jumlah responden menurut Crocker dan Algina menyatakan bahwa untuk memastikan kestabilan, setidaknya diperlukan 200 responden dalam hal jumlahnya dan menurut Nunnally (1970: 214-215), ukuran sampel yang ideal dalam uji coba adalah sepuluh kali lipat dari jumlah item yang terdapat dalam alat ukur tersebut [11]. Sehingga pada penelitian ini responden yang digunakan yaitu 218 responden.

### D. Instrumen Penelitian

Pada Instrumen penelitian ini ada beberapa indikator dan variabel yang disusun dari konstruk model UTAUT yang telah ditetapkan. Indikator dan variabel yang digunakan seperti yang terdapat pada Tabel I.

TABEL I  
 INSTRUMEN PENELITIAN

Variabel	Pernyataan	Indikator
Performance Expectancy	Saya merasa aplikasi CapCut memberikan banyak manfaat	PE01
	Saya merasa dengan menggunakan aplikasi CapCut membuat pekerjaan menjadi lebih efisien	PE02
	Saya merasa aplikasi CapCut meningkatkan produktifitas saya	PE03

Variabel	Pernyataan	Indikator
<i>Effort Expectancy</i>	Saya merasa aplikasi CapCut mudah digunakan dan dioperasikan	EE01
	Saya merasa aplikasi CapCut memiliki proses layanan yang tidak membingungkan/rumit	EE02
	Saya merasa fitur pada aplikasi CapCut mudah dipahami	EE03
	Saya tidak merasa kesulitan saat menggunakan aplikasi CapCut	EE04
<i>Social Influence</i>	Saya menggunakan aplikasi CapCut karena dipengaruhi oleh orang lain atau orang sekitar	SI01
	Kebanyakan orang di sekitar saya menggunakan aplikasi CapCut	SI02
	Banyak orang disekitar saya sangat mendukung saya dalam menggunakan aplikasi CapCut	SI03
	Saya merasa orang-orang di sekitar saya ataupun orang lain perlu menggunakan aplikasi CapCut	SI04
<i>Facilitating Conditions</i>	Saya merasa sangat penting menggunakan aplikasi CapCut karena nyaman digunakan	FC01
	Saya merasa sangat penting menggunakan aplikasi CapCut karena memudahkan pengguna	FC02
	Saya merasa aplikasi CapCut dapat memenuhi layanan editing	FC03
	Saya mendapatkan layanan bantuan dengan mudah ketika mengalami kendala	FC04

Variabel	Pernyataan	Indikator
<i>User Satisfaction</i>	saat menggunakan aplikasi CapCut	US
	Saya merasa puas dan berniat menggunakan aplikasi kembali	

Dari indikator dan variabel seperti yang di paparkan pada Tabel I. Terdapat kriteria dari setiap indikator-indikator tersebut dengan menggunakan skala Likert. Dimana Sekala Likert adalah alat pengukuran psikometri untuk mengukur sikap ataupun pendapat seseorang terhadap suatu topik dengan menggunakan skala kategori respons yang menunjukkan tingkat setuju atau tidak setuju [12].

Kriteria atau skala Likert yang akan diterapkan disajikan dalam Tabel II.

TABEL II  
SKALA LIKERT

Simbol	Kriteria Penelitian	Skor
SS	Sangat Setuju	5
S	Setuju	4
N	Netral	3
TS	Tidak Setuju	2
STS	Sangat Tidak Setuju	1

Dari kriteria penelitian yang disajikan pada Tabel II. Terdapat pengklasifikasian nilai dalam menentukan puas atau tidak puas seorang pelanggan tersebut. Dalam mengklasifikasikan nilai kepuasan pengguna atau user satisfaction menggunakan dua interval nilai seperti pada Tabel III [13]

TABEL III  
INTERVAL KLASIFIKASI LABEL

Interval	Klasifikasi
1-3	Tidak Puas
4-5	Puas

Pengklasifikasian kriteria penelitian ini seperti yang disajikan pada Tabel III Klasifikasi ini kemudian dijadikan label pada penelitian ini.

#### E. Data Cleaning

Dalam tahap data cleaning disini dilakukan pembersihan data agar data menjadi relevan untuk digunakan. Pembersihan data ini dilakukan untuk mengatasi jika ada data yang kosong, data terduplikasi, data noise, dan lain sebagainya.

#### F. Pemilihan Atribut

Pada tahap pemilihan atribut ini dilakukan dengan mengacu pada indikator pada variabel utama seperti yang disajikan pada Tabel 2. Atribut ini yang kemudian dijadikan sebagai bahan penelitian dalam mengetahui kepuasan pengguna aplikasi CapCut.

#### G. Pelabelan

Dalam memberikan label pada data, ini mengacu pada kriteria penelitian yang sudah diklasifikasikan menjadi 2 label yaitu label “Tidak Puas” dan label “Puas”. Dimana klasifikasi tersebut dipaparkan pada Tabel 4.

H. Uji Validitas dan Reliabilitas

Validitas adalah indikator kesahihan suatu instrumen yang menunjukkan kemampuannya dalam mengukur data dari variabel yang diteliti dengan tepat, di mana instrumen yang memiliki validitas tinggi dianggap sah, dan demikian pula sebaliknya [14]. Pengujian validitas instrumen pada penelitian ini merupakan suatu ketepatan dan kecermatan alat/instrumen penelitian dalam mengukur sesuatu yang ingin diukur dalam penelitian [10].

Adapun persamaan Koefisien Korelasi Produk Momen adalah sebagai berikut:

$$r_{cy} = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (1)$$

Keterangan:

- $r_{cy}$  = koefisien korelasi variabel X dan Y
- $n$  = jumlah responden
- $\sum X$  = jumlah skor butir pernyataan
- $\sum Y$  = jumlah skor total pernyataan

Dasar untuk menilai validitas adalah jika nilai r-hitung melampaui nilai r-tabel, menunjukkan korelasi signifikan antara item-item angket dan skor keseluruhan, sehingga dianggap valid. Namun jika nilai r-hitung tidak mencapai nilai r-tabel, maka item-item tersebut dianggap tidak valid karena kurangnya korelasi yang signifikan dengan skor keseluruhan [15].

Reliabilitas yaitu tingkat keakuratan, ketepatan atau ketelitian suatu instrumen, dimana instrumen tersebut konsisten memberikan hasil yang serupa dalam pengukuran suatu variabel pada berbagai waktu [14]. Uji reliabilitas memiliki tujuan untuk menilai konsistensi kuesioner dalam penelitian, dengan memastikan bahwa hasil pengukuran tetap sama meskipun dilakukan berulang kali [16].

Untuk menguji reliabilitas instrumen penelitian digunakan metode pengukuran dengan menggunakan rumus Cronbach’s Alpha. Berikut adalah persamaan Cronbach’s Alpha :

$$r_{11} = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right) \quad (2)$$

Keterangan :

- $r_{11}$  = reliabilitas kuesioner
- $k$  = jumlah item
- $\sum \sigma_b^2$  = jumlah varian item
- $\sigma_t^2$  = total varian

Sebuah instrumen dapat dikatakan reliabel apabila mempunyai koefisien Cronbach Alpha lebih dari 0,60 [6].

I. Principal Component Analysis (PCA)

Untuk mencari faktor yang paling berpengaruh dalam penelitian ini menggunakan metode *Principal Component Analysis* (PCA). Metode PCA merupakan Metode Pendekatan banyak digunakan oleh para ahli statistik dan peneliti kuantitatif [10]. Untuk mengetahui hasil dari ekstraksi yang lebih detail menggunakan metode PCA, dapat diketahui melalui uji total variance yang di dalamnya menunjukkan nilai eigen [4]. PCA merupakan salah satu teknik prosedur statistik yang dipakai oleh Kaisers criterion. Berdasarkan kriteria Kaiser, hanya faktor yang mempunyai nilai 1.0 maupun lebih yang perlu dianalisis atau juga disebut pembentuk faktor [10].

J. Modelling

Tahap modelling merupakan tahap pengaplikasian algoritma yang akan digunakan yaitu algoritma K- Nearest Neighbor (K-NN). Setelah dilakukan (PCA) data dapat digunakan untuk dimodelkan dengan menggunakan K- Nearest Neighbor, dalam penggunaan algoritma ini dilakukan penghitungan dengan menggunakan Persamaan berikut :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (3)$$

Keterangan :

- $d(x_i, x_j)$  : Jarak Euclidean (Euclidean Distance) (jarak antar data training dan data testing)
- $x_i$  : Record ke-i (data training)
- $x_j$  : Record ke-j (data testing)
- $a_r$  : data ke-r
- $i, j$  : 1,2,3,...n
- $n$  : banyak data

Dari persamaan tersebut ada beberapa langkah-langkah untuk menghitung algoritma K-NN, yaitu menetapkan nilai k, lalu kuadrat jarak Euclidean dihitung antara data uji dan data latih, dan diurutkan dari yang terkecil, kemudian label terbanyak dari tetangga terdekat diambil untuk menentukan klasifikasi data uji. [8].

K. Evaluasi

Pada tahap evaluasi ini dilakukan untuk melihat bagaimana kinerja atau performance dari metode dan algoritma yang digunakan. Metode evaluasi yang digunakan dalam penelitian ini adalah Confusion Matrix. Confusion matrix merupakan alat evaluasi untuk menilai performa suatu masalah klasifikasi yang memiliki dua kelas ataupun lebih sebagai keluarannya [17]. Terdapat empat kombinasi berbeda antara nilai prediksi dan nilai aktual yang terdapat dalam confusion matrix, yaitu true positives, true negatives, false positives, dan false negatives. Keempat metode evaluasi ini digunakan untuk menilai kinerja. Informasi mengenai akurasi juga dapat ditemukan dalam Tabel IV berikut ini yang menggunakan confusion matrix [18].

TABEL IIIV  
 CONFUSION MATRIX

	<i>Actual Class</i>
--	---------------------

		Yes	No
Predicted Class	Yes	Number of True Positive (TP)	Number of False Positive (FP)
	No	Number of False Negative (FN)	Number of True Negative (TN)

Keterangan:

1. TP (True Positive): Merupakan kasus di mana kelas yang diprediksi adalah Y dan kelas yang sebenarnya juga Y.
2. TN (True Negative): Merupakan kasus di mana kelas yang diprediksi adalah N dan kelas yang sebenarnya juga N.
3. FP (False Positive): Merupakan kasus di mana kelas yang diprediksi adalah Y tetapi kelas yang sebenarnya adalah N.
4. FN (False Negative): Merupakan kasus di mana kelas yang diprediksi adalah N tetapi kelas yang sebenarnya adalah Y.

Untuk performa dengan menggunakan Confusion Matrix, terdapat 3 nilai kombinasi untuk melakukan perhitungan Confusion Matrix, yaitu accuracy, precision, dan recall [19]. Untuk rumus perhitungan sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Keterangan :

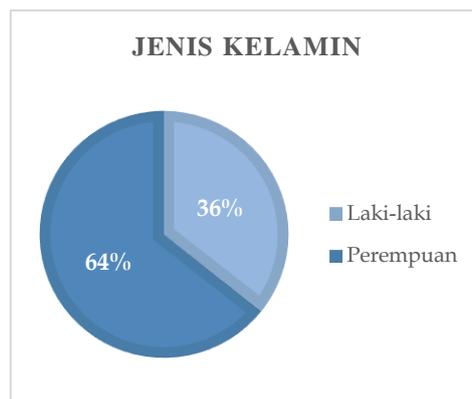
1. Akurasi adalah ukuran yang menunjukkan seberapa baik pengklasifikasi dapat memprediksi kelas dari setiap instance dengan tepat.
2. Precision mengacu pada sejauh mana pengklasifikasi dapat memprediksi kelas yang benar-benar relevan dan sesuai.
3. Recall menunjukkan sejauh mana pengklasifikasi dapat mendeteksi kelas dari instance yang relevan di antara semua instance yang relevan dalam kelas tersebut.

### III. HASIL DAN PEMBAHASAN

#### A. Hasil Pengumpulan Data

Pada tahap disini dilakukan analisis dari data yang terkumpul. Penelitian ini melibatkan 218 responden yang dikumpulkan dengan teknik penyebaran kuesioner dengan informasi demografis yang mencakup jenis kelamin, usia, lama penggunaan aplikasi CapCut, terakhir kali penggunaan aplikasi, dan domisili. Berikut adalah hasil analisis demografi responden :

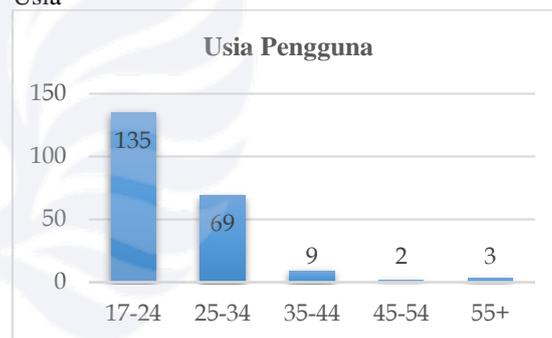
##### 1. Jenis Kelamin



Gbr. 3 Data Jenis Kelamin

Seperti yang ditampilkan pada Gbr. 3, data jenis kelamin menunjukkan bahwa dari 219 responden, 35,30% adalah laki-laki dan 64,70% adalah perempuan. Hal ini menunjukkan bahwa aplikasi CapCut lebih populer di kalangan perempuan. Tingginya persentase perempuan yang menggunakan aplikasi ini dapat diinterpretasikan sebagai indikasi bahwa fitur dan konten yang disediakan oleh CapCut lebih menarik bagi perempuan.

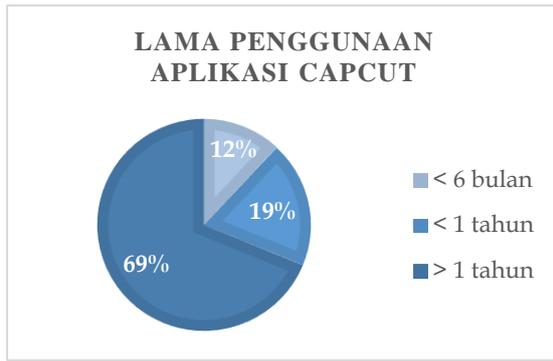
##### 2. Usia



Gbr. 4 Data Usia Pengguna

Seperti yang ditampilkan pada Gbr. 4, Responden penelitian ini didominasi oleh kelompok usia 17-24 tahun sebanyak 135 orang (61.83%), diikuti oleh usia 25-34 tahun sebanyak 69 orang (31.65%). Sementara itu, kelompok usia 35-44, 45-54, dan 55+ masing-masing memiliki 9 (4.13%), 2 (0.92%), dan 3 responden (1.38%). Hal ini menunjukkan bahwa mayoritas pengguna aplikasi CapCut adalah generasi muda. Generasi ini lebih tertarik untuk menggunakan aplikasi pengeditan video seperti CapCut untuk membuat dan membagikan konten kreatif mereka. Dengan tingginya minat dari kelompok usia muda, CapCut dapat terus mengembangkan fitur-fitur yang sesuai dengan tren dan kebutuhan pengguna muda untuk mempertahankan basis pengguna yang besar ini.

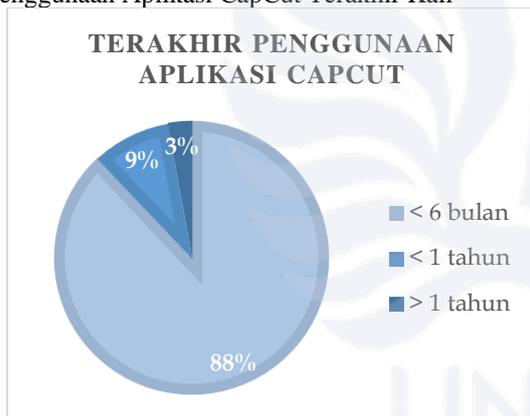
##### 3. Lama Pengguna Aplikasi CapCut



Gbr. 5 Data Lama Penggunaan Aplikasi CapCut

Seperti yang ditampilkan pada Gbr. 5, kebanyakan responden sudah menggunakan aplikasi layanan CapCut dengan lebih dari 1 tahun, yaitu sebanyak 68.8% (150 orang). Responden yang telah menggunakan aplikasi selama kurang dari 1 tahun sebanyak 19.3% (42 orang), dan yang menggunakan kurang dari 6 bulan sebanyak 11.9% (26 orang). Data ini menunjukkan bahwa sebagian besar pengguna adalah pengguna Lama. Hal ini bisa disebabkan oleh popularitas CapCut yang semakin meningkat.

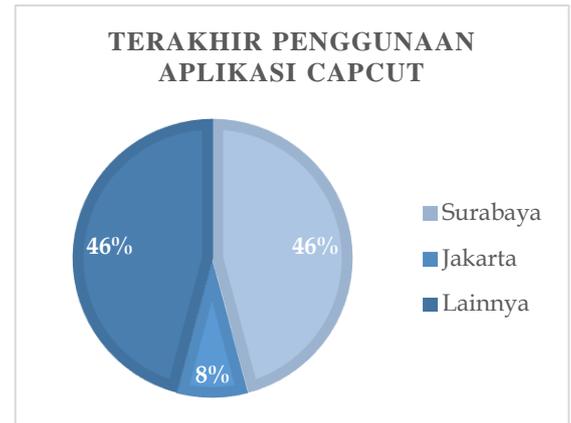
4. Penggunaan Aplikasi CapCut Terakhir Kali



Gbr. 6 Data Terakhir Penggunaan Aplikasi CapCut

Seperti yang sudah ditampilkan pada Gbr. 6, sebanyak 88.1% (192 orang) responden melaporkan bahwa mereka menggunakan aplikasi CapCut dalam 6 bulan terakhir. Sebanyak 9.2% (20 orang) menggunakan aplikasi ini di jangka waktu kurang dari 1 thn, dan 2.8% (6 orang) menggunakan lebih dari 1 tahun. Ini konsisten dengan temuan bahwa mayoritas pengguna adalah pengguna lama yang terus memanfaatkan aplikasi ini sebagai media kreatif mereka. Penggunaan dalam 6 bulan terakhir yang tinggi menunjukkan bahwa aplikasi CapCut berhasil mempertahankan minat pengguna.

5. Domisili

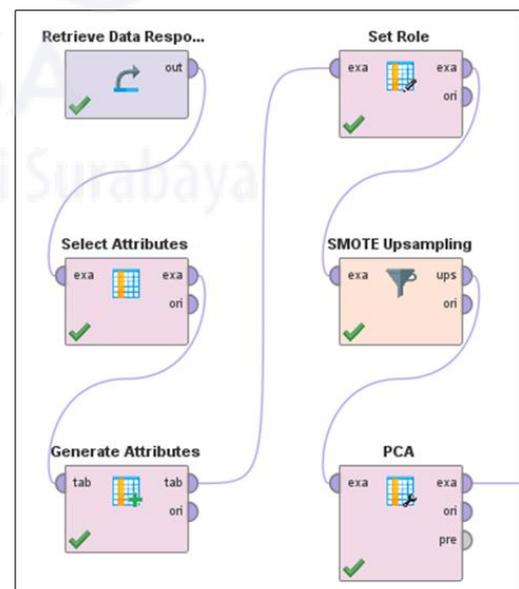


Gbr. 7 Data Terakhir Penggunaan Aplikasi CapCut

Seperti yang ditampilkan pada Gbr. 7, sebagian besar responden berasal dari Surabaya (45,87%, 100 orang), diikuti oleh Jakarta (8,26%, 18 orang), dan lainnya (45,87%, 100 orang). Hal ini menunjukkan bahwa pengguna CapCut tersebar di berbagai wilayah, namun lebih banyak terkonsentrasi di Surabaya. Pengguna dari wilayah lain yang cukup signifikan menunjukkan bahwa CapCut juga memiliki daya tarik yang luas dan diterima dengan baik di berbagai kota di Indonesia. Keberagaman lokasi pengguna ini menunjukkan potensi pasar yang luas bagi CapCut untuk mengembangkan strategi pemasaran dan pengembangan fitur yang lebih menyeluruh dan inklusif.

B. Hasil Penyortiran dan Labeling Data

Dalam penelitian ini didapatkan hasil jawaban kuesioner oleh responden sejumlah 218 data. Dan kemudian dilakukan preprocessing data seperti yang ditampilkan pada Gbr. 8.



Gbr. 8 Tahapan Data Preprocessing

Seperti yang dapat dilihat pada Gbr. 8 ada beberapa tahap preprocessing data, yaitu sebagai berikut :

a. Pemilihan atribut

Pememilihan atribut digunakan sebagai parameter pada model yang akan dibuat. Pemilihan atribut ini menggunakan operator "Attribute Selections" pada Rapid Miner. Atribut yang digunakan dalam pemilihan ini adalah No. Hp, PE01, PE02, PE03, EE01, EE02, EE03, EE04, , SI01, SI02, SI03, SI04, FC01, FC02, FC03, FC04, dan US.

b. Pelabelan

Kemudian dilakukan penambahan atribut pada data yang ada dengan menggunakan operator "Generate Attribute" pada Rapidminer. Atribut yang ditambahkan adalah atribut "Keterangan" yang akan digunakan sebagai label pada setiap data. Proses pemberian label berdasarkan nilai pada item User Satisfaction (US) pada setiap data yang ada dan mengacu pada interval yang telah ditentukan yaitu, label PUAS untuk data dengan nilai > 3, dan label TIDAK PUAS untuk data dengan nilai < 4. Pemberian status label ini menggunakan operator "Set role" pada Rapidminer dan hasil yang dapat dilihat seperti yang disajikan pada Tabel V dan Tabel VI.

TABEL V  
SEBELUM PROSES LABELING

PE01	PE02	...	US
4	5		5
4	4		5
...	...		...
3	3		2

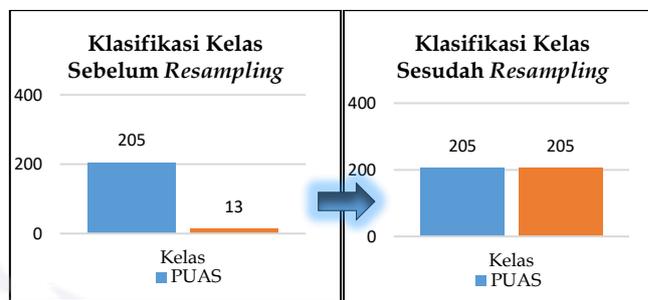
TABEL VI  
SETELAH PROSES LABELING

PE01	PE02	...	US	Keterangan
4	5		5	PUAS
4	4		5	PUAS
...	...		...	
3	3		2	TIDAK PUAS

Dari hasil pemberian label pada Tabel VI data kuesioner, diketahui sebanyak 205 data terklasifikasi PUAS dan 13 data terklasifikasi TIDAK PUAS.

C. Resampling

Dari hasil pengklasifikasian label tersebut ternyata data mengalami imbalanced data sehingga perlu dilakukan resampling menggunakan operator rapidminer "SMOTE Upsampling" dengan hasil seperti pada Gbr. 9.



Gbr. 9 Hasil Resampling Data

Setelah melakukan resampling menggunakan teknik Oversampling, diperoleh data yang telah seimbang yaitu seperti yang ditampilkan pada Gbr. 9, sehingga keseimbangan data tersebut memenuhi kategori normal dan siap dipakai untuk pemodelan.

D. Hasil Uji Validitas dan Reliabilitas Dataset

Tahap ini merupakan pengujian terkait validitas dan reliabilitas dari data kuesioner yang telah diseleksi. Hal ini untuk mengetahui apakah dataset tersebut dapat dikatakan valid dan memiliki keandalan. Data yang digunakan dalam uji instrumen ini berjumlah 218 data.

a. Uji Validitas

TABEL VIV  
HASIL UJI VALIDITAS

Indikator	Corrected Item-Total Correlation	Nilain R-Tabel	Keterangan
PE01	0,459	0,138	VALID
PE02	0,349	0,138	VALID
PE03	0,494	0,138	VALID
EE01	0,528	0,138	VALID
EE02	0,578	0,138	VALID
EE03	0,633	0,138	VALID
EE04	0,607	0,138	VALID
SI01	0,537	0,138	VALID
SI02	0,691	0,138	VALID
SI03	0,687	0,138	VALID
SI04	0,686	0,138	VALID
FC01	0,771	0,138	VALID
FC02	0,762	0,138	VALID

Indikator	Corrected Item-Total Correlation	Nilain R-Tabel	Keterangan
FC03	0,658	0,138	VALID
FC04	0,603	0,138	VALID

Berdasarkan Tabel VII, dapat diambil kesimpulan bahwa keseluruhan data dan instrumen penelitian untuk variabel yang digunakan dalam studi ini dpt dinyatakan valid karena mempunyai nilai  $r$  hitung melebihi 0,138.

b. Uji Reliabilitas

Reliabilitas kuesioner diukur menggunakan rumus Cronbach's Alpha, dan sebuah instrumen dianggap reliabel apabila mempunyai koefisien Cronbach Alpha melebihi 0,60.

TABEL VII  
HASIL UJI RELIABILITAS PERFORMANCE EXPECTANCY (PE)

Kriteria Pengujian Performance Expectancy		
Nilai Acuan	Nilai Chronbach's Alpha	Keterangan
0,6	0,603	RELIABEL

Berdasarkan uji reliabilitas yang ditampilkan pada Tabel VIII, diperoleh nilai Cronbach's Alpha sebesar 0,603, yang melebihi batas acuan 0,60. Dengan ini, dapat diambil kesimpulan bahwa item pernyataan yang digunakan adalah reliabel.

TABEL IX  
HASIL UJI RELIABILITAS EFFORT EXPECTANCY (EE)

Kriteria Pengujian Effort Expectancy		
Nilai Acuan	Nilai Chronbach's Alpha	Keterangan
0,6	0,714	RELIABEL

Berdasarkan uji reliabilitas yang telah dilakukan Berdasarkan uji reliabilitas yang ditunjukkan pada Tabel IX, diperoleh nilai Cronbach's Alpha sebesar 0,714, yang melebihi nilai acuan 0,60. Dengan hasil tersebut, item pernyataan yang digunakan dapat dianggap reliabel.

TABEL X  
HASIL UJI RELIABILITAS SOCIAL INFLUENCE (SI)

Kriteria Pengujian Social Influence		
Nilai Acuan	Nilai Chronbach's Alpha	Keterangan
0,6	0,766	RELIABEL

Berdasarkan uji reliabilitas yang ditunjukkan pada Tabel X, didapatkan nilai Cronbach's Alpha sebesar 0,766, yang melebihi batas acuan 0,60. Oleh karena itu, item pernyataan yang digunakan dapat dianggap reliabel.

TABEL XI  
HASIL UJI RELIABILITAS FACILITATING CONDITIONS (FC)

Kriteria Pengujian Facilitating Conditions		
Nilai Acuan	Nilai Chronbach's Alpha	Keterangan
0,6	0,824	RELIABEL

Berdasarkan uji reliabilitas yang ditunjukkan pada Tabel XI, diperoleh nilai Cronbach's Alpha sebesar 0,824, yang melampaui batas acuan 0,60. Dengan demikian, item pernyataan yang digunakan dapat dinyatakan reliabel.

E. Hasil Principal Component Analysis (PCA)

- Menggunakan perhitungan manual  
Menghitung Standarisasi Data Hitung mean ( $\mu$ ) dan standar deviasi ( $\sigma$ ) untuk setiap variabel:

$$\mu_{PE01} = \frac{5+5+5+4+\dots+5}{218} = \frac{974}{218} = 4,47$$

$$\sigma_{PE01} = \sqrt{\frac{(5-4,47)^2+(5-4,47)^2+\dots+(5-4,47)^2}{218}}$$

$$\sqrt{\frac{78,2762}{218}} = \sqrt{0,359065138} = 0,6$$

Mencari Standarisasi data :

$$z_{ii} = \frac{n - \mu}{\sigma}$$

Keterangan:

$z_{ii}$  = Standarisasi data

$n$  = nilai sampel

$\mu$  = Mean (rata-rata)

$\sigma$  = Standar deviasi

$$z_{11} = \frac{5 - 4,47}{0,6} = 0,9$$

Mencari matriks kovariansi:

$$cov(z_i, z_j) = \frac{1}{N-1} \sum_{i=1}^N (z_{ii} \times z_{jj})$$

Keterangan:

$z_i$  = variabel ke-i

$z_j$  = variabel ke-j

$N$  = jumlah sampel

$$cov(PE01, PE02) = \frac{1}{218-1} ((0,9 \times (-0,5)) + ((0,9 \times 1) + (0,9 \times (-0,5)) + \dots + (0,9 \times 1)) = 0,28$$

$$cov(PE01, PE03) = \frac{1}{218-1} ((0,9 \times (-1,6)) + (0,9 \times 1) + (0,9 \times (-0,3)) + \dots + (0,9 \times (-0,3)) = 0,4$$

$$cov(PE01, EE01) = \frac{1}{218-1} ((0,9 \times (-0,4)) + (0,9 \times 9) + (0,9 \times (-0,9)) + \dots + (0,9 \times (-0,4)) = 0,41$$

Kalkulasi kovariansi dengan diri sendiri:

$$cov(PE01, PE01) = \frac{1}{217} \sum_{i=1}^N (z_{PE01} \times z_{PE01}) = 1.00$$

$$cov(PE02, PE02) = \frac{1}{217} \sum_{i=1}^N (z_{PE02} \times z_{PE02}) = 1.00$$

Dari perhitungan diatas secara akumulasi didapatkan hasil matriks kovariansi sebagai berikut :

$$A = \begin{pmatrix} 1 & \dots & 0.3260 \\ \vdots & \ddots & \vdots \\ 0.3260 & \dots & 1 \end{pmatrix}$$

Selanjutnya Mencari eigenvalue

$$A - \lambda I = \begin{pmatrix} 1 - \lambda & 0.2777 & \dots & 0.3260 \\ 0.2777 & 1 - \lambda & \dots & 0.2085 \\ \vdots & \vdots & \ddots & \vdots \\ 0.3260 & 0.2085 & \dots & 1 - \lambda \end{pmatrix}$$

$$\det(A - \lambda I) = 0$$

Keterangan:

A = Hasil matriks kovariansi

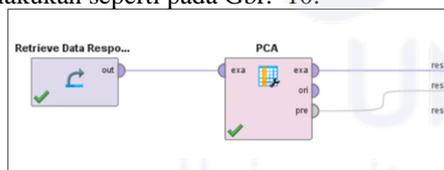
$\lambda$  = eigenvalue

Didapatkan nilai eigenvalue

$$\begin{aligned} \lambda_1 &= 6.05498979, \lambda_2 = 1.74501083, \lambda_3 \\ &= 0.95529133, \lambda_4 = 0.93687853, \\ \lambda_5 &= 0.83690635, \lambda_6 \\ &= 0.78429468, \lambda_7 = 0.71466575, \lambda_8 \\ &= 0.23902082, \lambda_9 = 0.62134377, \lambda_{10} \\ &= 0.58605561, \lambda_{11} \\ &= 0.30943741, \lambda_{12} \\ &= 0.52508241, \lambda_{13} \\ &= 0.49373967, \lambda_{14} = 0.4444356, \lambda_{15} \\ &= 0.36324592, \lambda_{16} = 0.38960153 \end{aligned}$$

## 2. Menggunakan Rapidminer

Hasil yang didapatkan dengan menggunakan metode (PCA) didapatkan faktor yang paling berpengaruh pada penelitian ini. Penggunaan metode ini dilakukan seperti pada Gbr. 10.



Gbr. 10 Proses Metode PCA

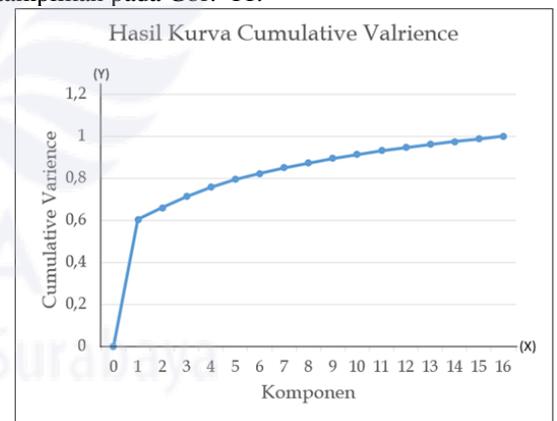
Dari proses dengan menggunakan metode PCA seperti yang disajikan pada Gbr. 10, didapatkan hasil cumulative variance seperti pada Tabel XII dengan 16 komponen utama. Cumulative variance ini menjelaskan bagaimana persentase komponen dapat menjelaskan keseluruhan data.

TABEL XII  
 HASIL CUMULATIVE VARIANCE PCA

Komponen	Cumulative Variance
1	0.604

Komponen	Cumulative Variance
2	0.659
3	0.713
4	0.757
5	0.794
6	0.823
7	0.850
8	0.873
9	0.894
10	0.914
11	0.931
12	0.947
13	0.962
14	0.975
15	0.988
16	1.000

Dari hasil cumulative variance yang disajikan pada Tabel XII, dapat terbentuk kurva seperti yang ditampilkan pada Gbr. 11.



Gbr. 11 Hasil Kurva Cumulative Variance

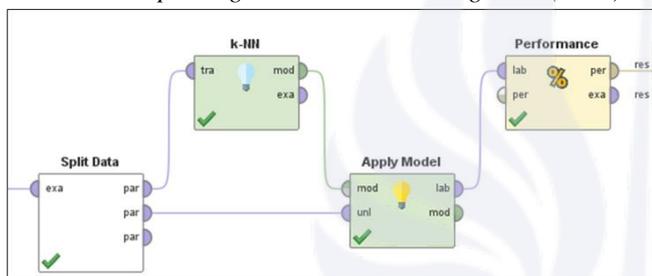
Kurva yang ditunjukkan pada Gbr. 11, menunjukkan bahwa jumlah kumulatif dari variance ratio meningkat seiring bertambahnya jumlah komponen. Kumulatif variance ratio mencerminkan total gabungan nilai variance ratio yang menjelaskan variabel-variabel tersebut. Hasil PCA mengidentifikasi variabel yang mempengaruhi kepuasan pengguna layanan aplikasi CapCut, dengan nilai di atas 0,5 seperti yang tercantum dalam Tabel XIII.

TABEL XIII  
VARIABEL YANG BERPENGARUH TERHADAP KEPUASAN PENGGUNA  
LAYANAN APLIKASI CAPCUT

No	Atribut	Komponen	Nilai
1	EE02	PC10	0,654
2	EE01	PC7	0,618
3	FC01	PC 9	0,599
4	FC01	PC16	0,549

Tabel XIII mengidentifikasi faktor-faktor yang mempengaruhi kepuasan pengguna layanan aplikasi CapCut, yaitu EE02, EE01, dan FC01. Variabel-variabel ini memiliki nilai di atas 0,5, yang menunjukkan kemampuannya dalam menjelaskan keseluruhan data [20].

#### F. Hasil Penerapan Algoritma k-Nearest Neighbors (KNN)



Gbr. 12 Proses Modelling Menggunakan kNN

Penerapan algoritma K-Nearest Neighbors (KNN) dilakukan dengan beberapa tahapan seperti yang disajikan pada Gbr. 12. Hasil yang didapatkan pada penerapan algoritma K-Nearest Neighbors (KNN) sebagai berikut :

##### 1. Split Data

Sebelum dilakukan modelling, dilakukan tahap membagi dataset menjadi dua yaitu data train (pelatihan) dan data test (pengujian). Data pelatihan akan digunakan untuk melatih model, sedangkan data pengujian untuk melakukan pengujian terhadap seberapa akurat model tersebut dapat memprediksi hasil yang benar. Untuk mendapatkan hasil yang efektif dilakukan perbandingan dalam train-test-split data dengan perbandingan 70%:30%, 80%:30%, dan 90%:10%. Sehingga didapatkan pembagian dataset , yaitu:

- 80% data digunakan sebagai data pelatihan.
- 20% data digunakan sebagai data pengujian.

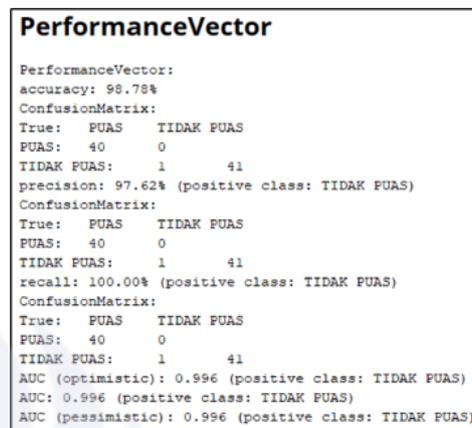
##### 2. KNN

Pada metode K-NN, nilai K digunakan untuk menentukan jumlah tetangga terdekat dari data latih. Pemilihan nilai K didasarkan pada akurasi hasil yang diperoleh. Untuk menentukan nilai K, dilakukan perbandingan dari beberapa nilai K, mulai dari K=1 hingga K=21. Dalam penelitian ini, nilai K yang

dipakai adalah K=5 karena nilai tersebut menghasilkan akurasi tertinggi.

##### 3. Evaluasi

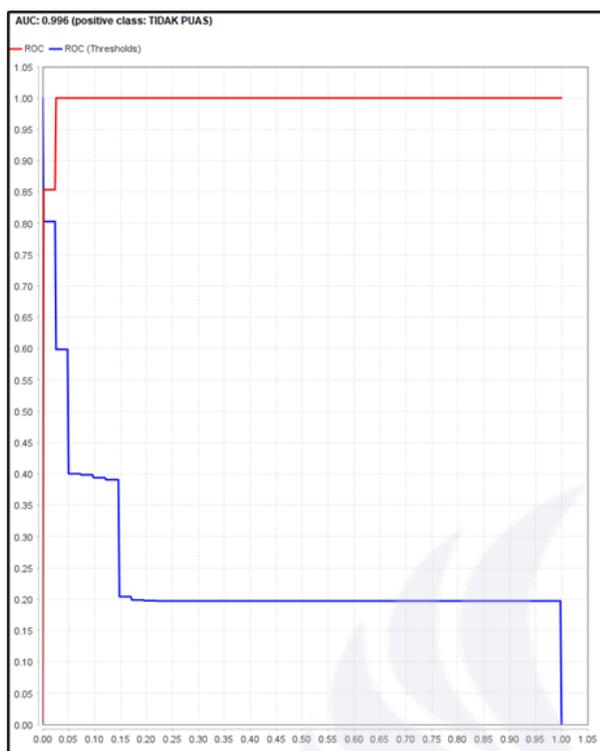
Pemodelan dengan menggunakan algoritma K-Nearest Neighbors (KNN) didapatkan hasil evaluasi kinerja seperti yang disajikan pada Gbr. 13.



Gbr. 13 Hasil Modelling Algoritma KNN

Dari Gbr. 13, didapatkan Hasil dari modeling menggunakan algoritma KNN dapat dijelaskan sebagai berikut:

- Akurasi: Model KNN memiliki akurasi sebesar 98,78%. Akurasi mengukur seberapa baik model dapat mengklasifikasikan secara benar baik kelas PUAS maupun TIDAK PUAS.
- Confusion Matrix:
  - True Positives (TP) untuk kelas TIDAK PUAS: Ada 41 prediksi yang benar untuk kelas TIDAK PUAS.
  - False Negatives (FN) untuk kelas TIDAK PUAS: Ada 1 prediksi yang salah untuk kelas TIDAK PUAS.
  - True Negatives (TN) untuk kelas PUAS: Ada 40 prediksi yang benar untuk kelas PUAS.
  - False Positives (FP) untuk kelas PUAS: Ada 0 prediksi yang salah untuk kelas PUAS.
- Precision: Untuk kelas TIDAK PUAS, nilai precision mencapai 97,62%. Ini berarti dari seluruh prediksi yang model berikan sebagai kelas TIDAK PUAS, semuanya benar.
- Recall (Sensitivity): Recall untuk kelas yg TIDAK PUAS adalah 100%. Ini menunjukkan bahwa dari semua instance yang sebenarnya TIDAK PUAS.



Gbr. 14 Hasil Area Under Curve (AUC)

Selain itu didapatkan juga hasil Area Under Curve (AUC) seperti yang disajikan pada Gbr. 14, dimana dijelaskan bahwa didapatkan nilai AUC 0.996. Ini adalah pengukuran yang baik untuk seberapa baik model dapat membedakan antara kelas TIDAK PUAS dan PUAS.

#### IV. SIMPULAN DAN SARAN

##### A. Simpulan

Berdasarkan hasil penelitian yang telah dilakukan dengan menggunakan metode K-Nearest Neighbors (KNN), ditemukan kesimpulan sebagai berikut:

1. Dalam penelitian ini, dengan menggunakan PCA 16 komponen didapatkan faktor yang paling mempengaruhi kepuasan pengguna layanan aplikasi CapCut yaitu Effort Expectancy dengan fokus yaitu "Proses layanan yang tidak rumit", dimana 24 responden atau 11,01% responden merasa TIDAK PUAS dan 194 responden atau 88,99% responden merasa PUAS.
2. Dengan menggunakan metode K-Nearest Neighbors (KNN) dalam penelitian ini mendapatkan hasil akurasi sebesar 0,9878 atau 98,78%. Dan memiliki presisi sebesar 0,9762 atau 97,62%, dan recall 1,00 atau 100%.

Dengan demikian, dapat disimpulkan bahwa penelitian ini berhasil dalam mengidentifikasi faktor-faktor yang berpengaruh pada kepuasan pengguna layanan aplikasi CapCut dan menghasilkan model prediksi yang akurat.

##### B. Saran

Berdasarkan hasil dan pembahasan dari penelitian yang telah diuraikan, berikut adalah saran untuk pengembangan pada penelitian selanjutnya:

1. Melakukan pengujian dengan metode klasifikasi lainnya seperti Support Vector Machine, Logistic Regression, Decision Tree, dan K Nearest Neighbors sehingga dapat diperoleh model klasifikasi yang memiliki tingkat akurasi lebih tinggi dalam mengukur kepuasan pengguna aplikasi CapCut.
2. Menggunakan parameter yang didasarkan pada konstruk lain seperti EUCS (End User Computing Satisfaction) sehingga dapat diperoleh akurasi yang lebih tinggi.

#### REFERENSI

- [1] 2stallions.com. (2024). Content Creation Trends For 2024. Diakses pada 3 Mei 2024, dari <https://2stallions.com/blog/content-creation-trends-for-2024/>
- [2] Muhammad, N. (2024). Tiktok, Aplikasi yang Paling Banyak Diunduh di Indonesia pada 2023. Diakses pada 4 Mei 2024, dari <https://databoks.katadata.co.id/datapublish/2024/02/27/tiktok-aplikasi-yang-paling-banyak-diunduh-di-indonesia-pada-2023>
- [3] Rohmah, M. (2024). Apa itu CapCut. Diakses pada 20 Mei 2024, dari <https://dibimbing.id/blog/detail/apa-itu-CapCut-fitur-contoh-penggunaannya>
- [4] Wijayanti, M. A. K., Persada, S. F., & Nareswari, N. (2021). Analisis Faktor Kepuasan Pelanggan terhadap Layanan Perusahaan Daerah Air Minum. *JURNAL SAINS DAN SENI ITS*, 10(1), 47–52.
- [5] Wibowo, E. P. E., & Prapanca, A. (2022). Analisis User Experience Terhadap Tingkat Kepuasan Pelanggan pada Aplikasi Mubeat. *Jurnal Education and Development*, 10(2), 596–602
- [6] Chlrasasthi, Y., & Nuryana, I. K. D. (2023). Komparasi Algoritma Naïve Bayes dan Random Forest Dalam Mengetahui Kepuasan Pengguna Aplikasi Jenius. *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, 4(3), 64–72
- [7] Tapak, L., Abbasi, H., & Mirhashemi, H. (2019). Assessment of factors affecting tourism satisfaction using K-nearest neighborhood and random forest models. *BMC Research Notes*, 12(1).
- [8] Diansyah, S. (2022). Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN). *Jurnal Sistem Informasi Dan Teknologi*, 4(1), 7–12.
- [9] Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Sistem Komputer Dan Kecerdasan Buatan (SISKOM-KB)*, 4(1), 15–21.
- [10] Budiastuti, D., & Bandur, A. (2018). Validitas dan Reliabilitas Penelitian (1st ed.). Mitra Wacana Media.
- [11] Alwi, I. (2015). Kriteria Empirik dalam Menentukan Ukuran Sampel pada Pengujian Hipotesis Statistika dan Analisis Butir. *Formatif Jurnal Ilmiah Pendidikan MIPA*, 2(2), 140–148.
- [12] Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, 7(4), 396–403.
- [13] Chrishariyani, C. D. A. A. P., Rahman, Y., & Aini, Q. (2022). Kepuasan Pengguna Layanan Shopee Food Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Informasi Bisnis*, 2, 99–106.
- [14] Priatna, T. (2017). *Prosedur Penelitian Pendidikan* (Nurhamzah, Ed.). CV. Insan Mandiri.
- [15] Al Hakim, R., Mustika, I., & Yuliani, W. (2021). Validitas dan Reliabilitas Angket Motivasi Berprestasi. *FOKUS (Kajian Bimbingan & Konseling Dalam Pendidikan)*, 4(4), 263–268.

- [16] Patmalasari, D., & Indriyanti, A. D. (2021). Analisis Kepuasan Pengguna Layanan Aplikasi MyTelkomsel dengan Menggunakan Model UTAUT. *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, 02(02), 37–45.
- [17] Febrian, A. K., Chrisnanto, Y. H., & Sabrina, P. N. (2022). Studi Komparasi Metode Klasifikasi K-Nearest Neighbor dan Naive Bayes dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk. *SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika*, 333–338.
- [18] Christian, J., Ernawati, I., & Chamidah, N. (2022). Implementasi Penggunaan Algoritma Categorical Boosting (Catboost) Dengan Optimisasi Hiperparameter Dalam Memprediksi Pembatalan Pesanan Kamar Hotel.
- [19] Salsabilah, B. N., & Nuryana, I. K. D. (2023). Komparasi Algoritma Naive Bayes dan K Nearest Neighbor dalam Kepuasan Pengguna Fitur Tiktok Shop. *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, 4(3), 31–39.
- [20] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374.

