

SENTIMENT ANALYSIS OF WONDR BY BNI APPLICATION EVALUATION USING MODIFIED K-NEAREST NEIGHBOR (MK-NN)

Arifatul Fathinah Essa
NIM. B2A223004

First Advisor : Tiani Wahyu Utami, M.Si
Second Advisor : Prizka Rismawati Arum, S.Si., M.Stat



ANALISIS SENTIMEN PENILAIAN APLIKASI WONDR BY BNI MENGGUNAKAN MODIFIED K-NEAREST NEIGBOR (MK-NN)

Arifatul Fathinah Essa
NIM. B2A223004

Pembimbing 1 : Tiani Wahyu Utami, M.Si
Pembimbing 2 : Prizka Rismawati Arum, S.Si., M.Stat



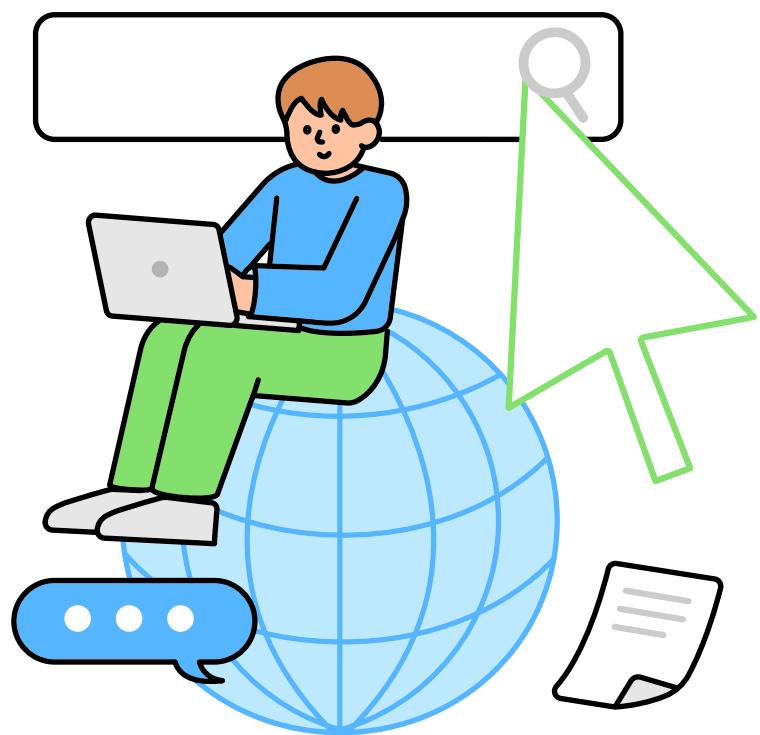
PENDAHULUAN

01.

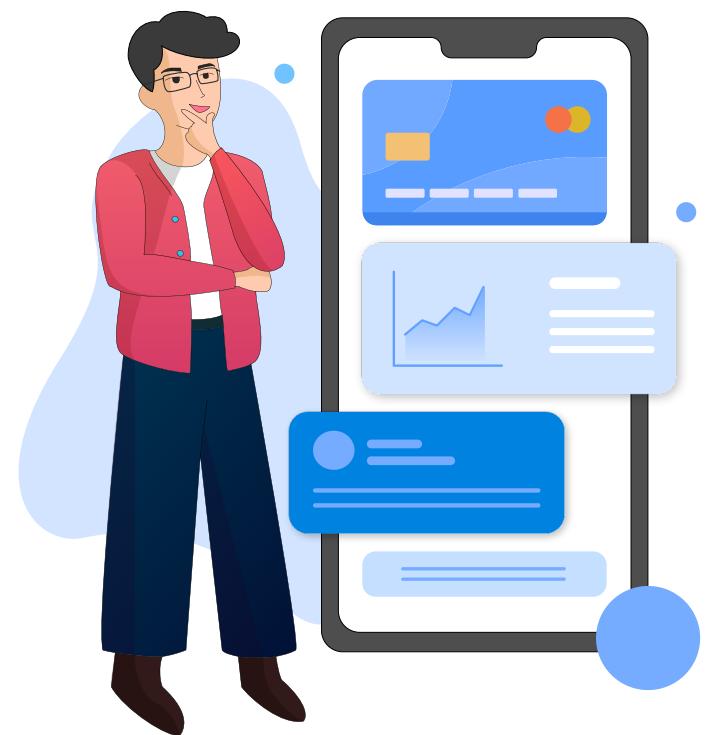
Latar Belakang



teknologi



internet



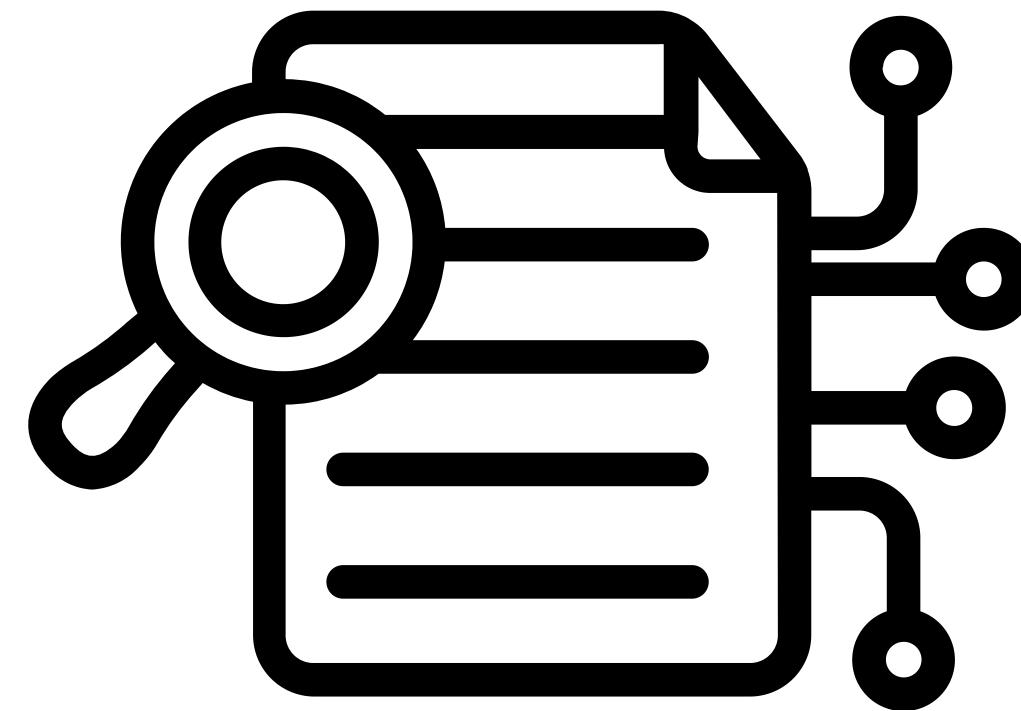
financial technology



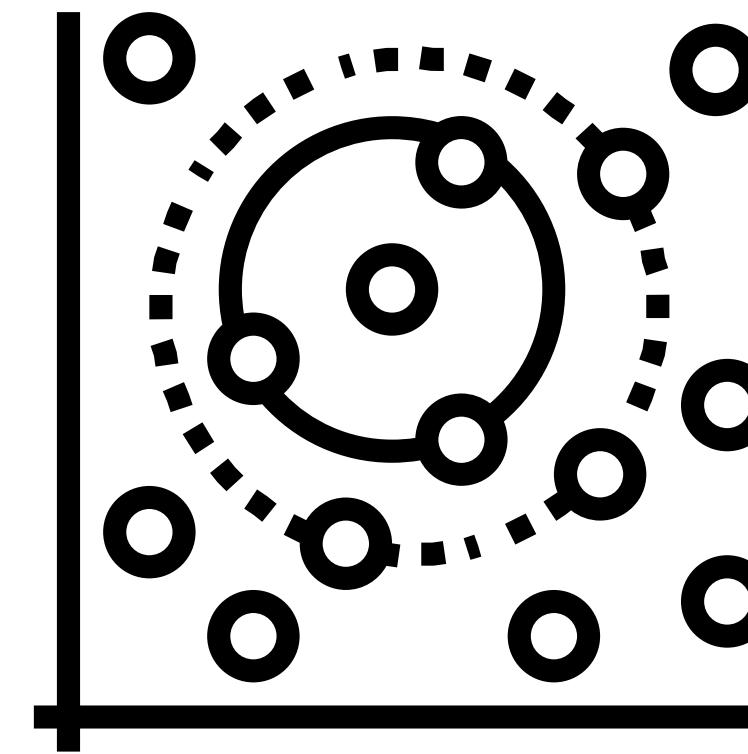
Latar Belakang



Latar Belakang



Text mining



K-Nearest Neighbor

+ Validity
+ weight voting

Modified K-Nearest Neighbor

Penelitian Terdahulu

1. Sofiah, dkk (2023)

Analisis sentimen kenaikan harga BBM menggunakan aplikasi Twitter.
Nilai akurasi 83,33%.
Nilai K yang digunakan adalah 3.

2. Paramitha, dkk (2020)

Ulasan pengguna MRT Jakarta.
Nilai akurasi 86%.
Nilai K yang digunakan adalah 3.

3. Halim, dkk (2023)

Prabowo Subianto bakal calon presiden 2024 pada aplikasi Twitter.
Nilai akurasi 93,3%.
Nilai K yang digunakan adalah 3.

Rumusan Masalah

Pertama

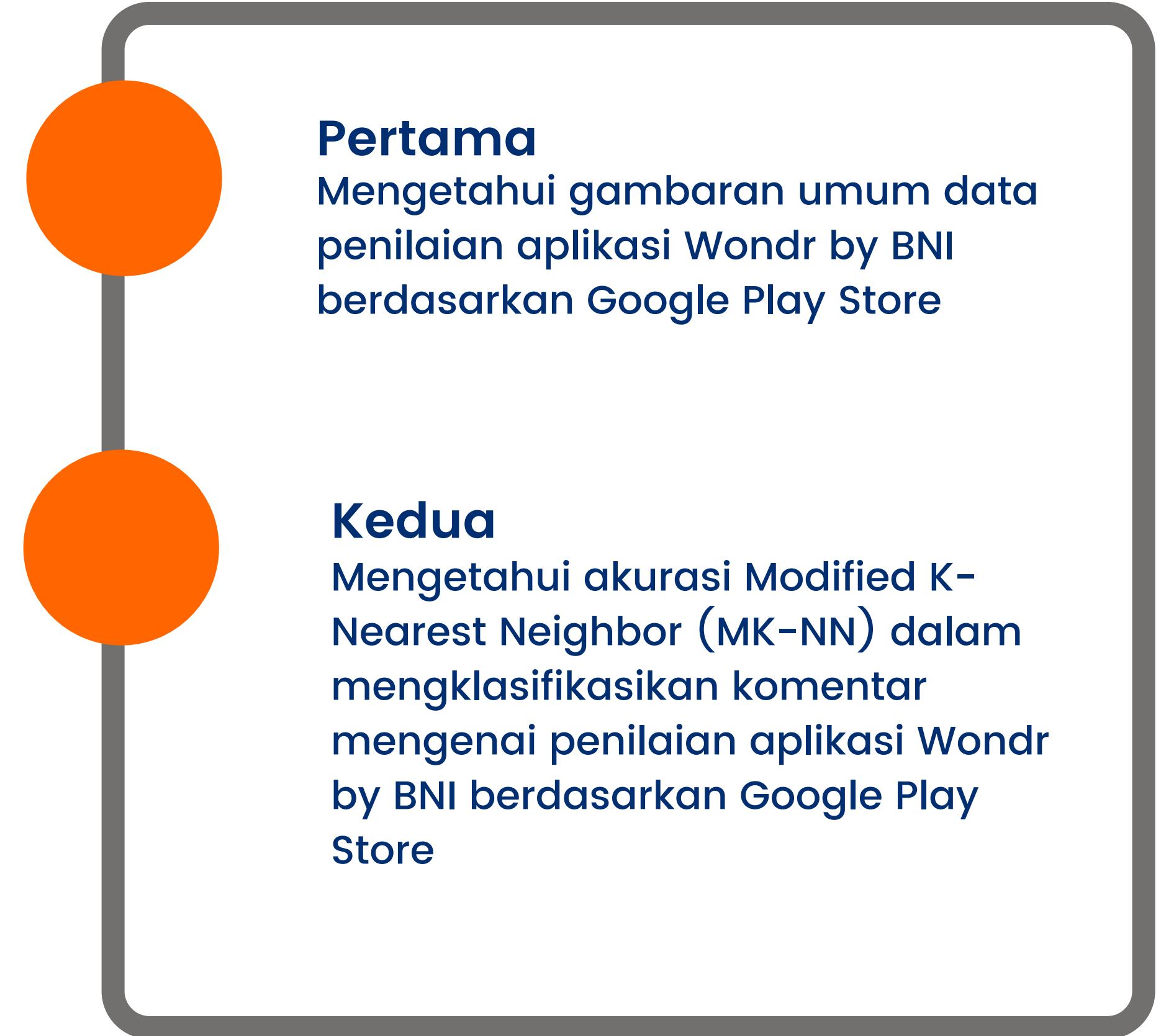
Bagaimana gambaran umum data penelitian aplikasi Wondr by BNI berdasarkan Google Play Store?



Kedua

Bagaimana akurasi Modified K-Nearest Neighbor dalam mengklasifikasikan komentar penilaian aplikasi Wondr by BNI berdasarkan Google Play Store?

TUJUAN PENELITIAN



Manfaat Penelitian

Bagi Pembaca

Memperluas pemahaman tentang penerapan statistika dalam pengklasifikasian text mining menggunakan Modified K-Nearest Neighbors (MK-NN)

Bagi Pembaca

Memungkinkan pembaca untuk memperdalam pengetahuan mereka tentang text mining menggunakan Modified K-Nearest Neighbors (MK-NN)

Bagi Universitas Muhammadiyah Semarang

Menjadi referensi yang berguna dalam mengembangkan pengetahuan tentang aplikasi statistika dalam konteks klasifikasi text mining.

Bagi Developer

Memberikan panduan praktis dalam mengetahui pendapat pengguna aplikasi baik dalam bentuk positif maupun negatif, sehingga dapat digunakan sebagai referensi untuk menjaga kualitas aplikasi

BATASAN PENELITIAN

1

Data yang digunakan merupakan data penilaian user aplikasi wondr by BNI yang terdapat di Google Play Store.

2

Analisis data akan dilakukan dengan bantuan perangkat lunak Python.

3

Jarak yang digunakan adalah cosine distance.

4

Data komentar yang digunakan dari 5 Juli sampai 2 september 2024

METODOLOGI PENELITIAN

02

METODOLOGI PENELITIAN

SUMBER DATA

Merupakan data sekunder yang diperoleh melalui proses scraping dari Google Play Store dengan data yang diperoleh sebanyak 4.264 data dari tanggal 5 Juli hingga 2 September 2024

VARIABEL PENELITIAN

- 1.Tanggal
- 2.Skor/bintang
- 3.Komentar



Tahapan penelitian

1

scraping data

scraping data pada Google Play Store dengan python

2

pre-processing data

data cleaning, case folding, tokenizing, stopword removal, stemming

3

pelabelan data

menggunakan bintang yang diberikan oleh user

4

visualisasi dengan wordcloud

visualisasi kata yang sering dibahas baik dari sentimen positif maupun negatif

5

Pembobotan kata

menggunakan TF-IDf

6

membagi data

membagi data menjadi data latih 80% dan data uji 20%

7

menghitung cosine similarity dan validitas

perhitungan antar data latih

8

menghitung cosine distance

perhitungan antara data latih dan data uji



Tahapan penelitian

9

menghitung nilai weight voting
sebagai penentu kelas data uji

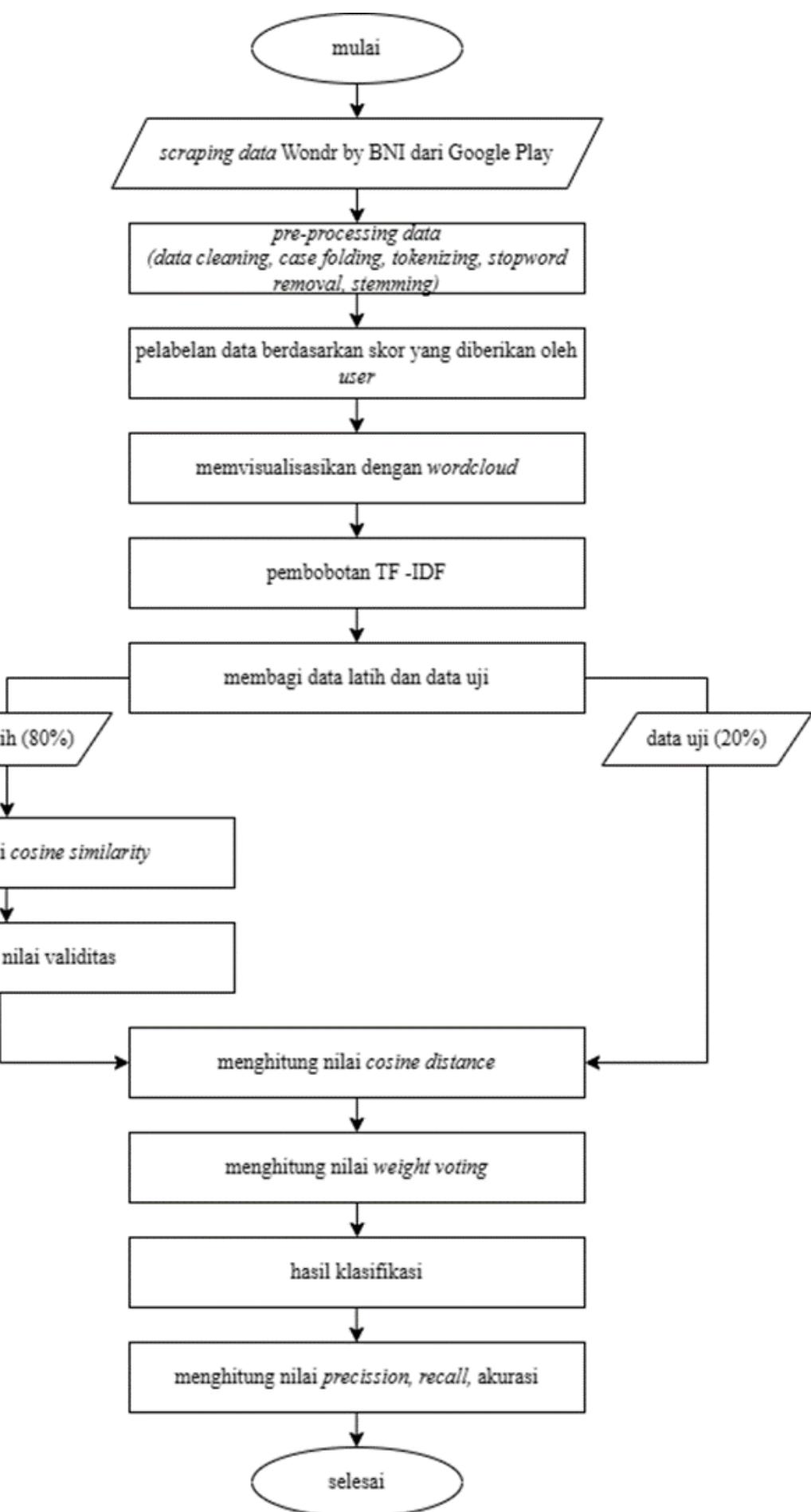
10

hasil klasifikasi
disajikan dalam confusion matrix

11

evaluasi model
menggunakan akurasi, presisi, dan recall

DIAGRAM ALIR PENELITIAN



HASIL DAN PEMBAHASAN

03

Scraping data

Data yang diperoleh dari tanggal 5 Juli hingga 2 September 2024 adalah sebanyak 4.264 data.

Pre-processing data

sebagai contoh, dipilih satu komentar random untuk tahapan pre-processing data yaitu:
“Makin mudah dan cepat bertransaksi dengan wondr  Wonder...  

1

2

3

4

5

Cleaning data

“Makin mudah dan cepat bertransaksi dengan wondr Wonder is wonderful”

Case folding

“makin mudah dan cepat bertransaksi dengan wondr wondr is wonderful”

Tokenizing

['makin', 'mudah', 'dan', 'cepat',
'bertransaksi', 'dengan', 'wondr', 'wondr',
'is', 'wonderful']

Stopword removal

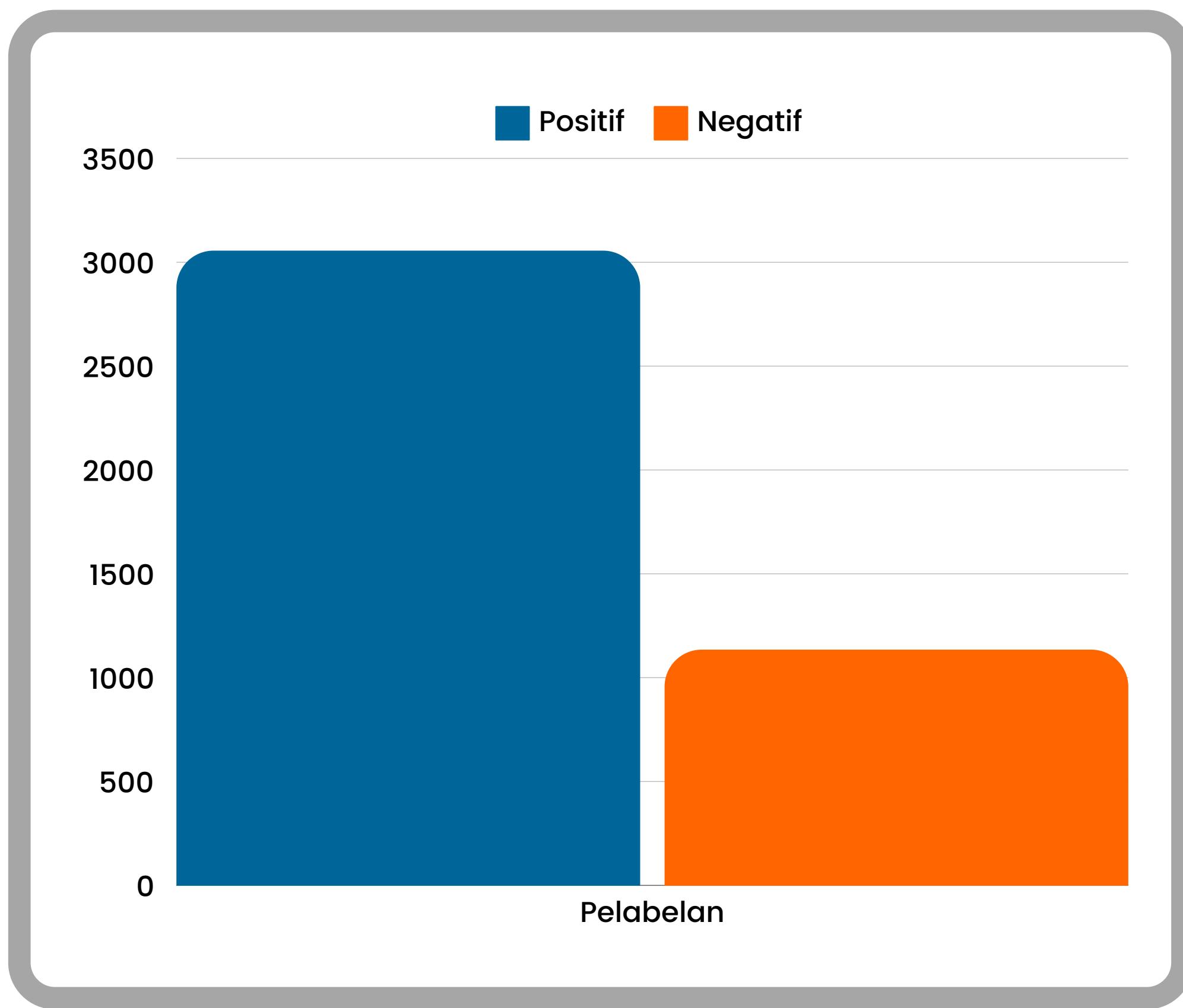
['mudah', 'cepat', 'bertransaksi', 'wondr',
'wondr', 'is', 'wonderful']

Stemming

['mudah', 'cepat', 'transaksi', 'wondr',
'wondr', 'is', 'wonderful']

Pelabelan

Dari komentar yang diberikan user, terdapat 3.056 sentimen positif dan 1.135 sentimen negatif



Wordcloud

Wordcloud positif



Wordcloud negatif



PEMBOBOTAN KATA DENGAN TF-IDF

Contoh menggunakan
10 komentar

D	Komentar	Label
D1	bayar qrис wondr cepat banget cuman 3 step selesai	Positif
D2	aplikasi top bayar qrис wondr cepat banget 3 langkah selesai mantap	Positif
D3	bayar pakai qrис cepat tampil mantap	Positif
D4	wondr simple bayar qrис cepat mobile banking cepat	Positif
D5	bayar qrис wondr cepat banget	Positif
D6	transaksi mudah qrис ewallet	Positif
D7	bayar qrис wondr cepat mudah jadi andal	Positif
D8	simple qrис rencana uang mudah simple	Positif
D9	kemarin top up voucher game pakai qrис aplikasi cepat banget tampil fresh banget warna bagus	Positif
D10	aplikasi aneh masak udah daftar udah transaksi daftar qrисnya gak transaksi	Negatif

PEMBOBOTAN KATA DENGAN TF-IDF

Hasil pembobotan dari contoh komentar

term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
andal	0	0	0	0	0	0	0.53	0	0	0
aneh	0	0	0	0	0	0	0	0	0	0.25
aplikasi	0	0.32	0	0	0	0	0	0	0.22	0.19
bagus	0	0	0	0	0	0	0	0	0.30	0
banget	0.32	0.28	0	0	0.54	0	0	0	0.40	0
banking	0	0	0	0.47	0	0	0	0	0	0
bayar	0.26	0.23	0.27	0.25	0.44	0	0.28	0	0	0
cepat	0.24	0.21	0.25	0.46	0.40	0	0.26	0	0.15	0
cuman	0.49	0	0	0	0	0	0	0	0	0
daftar	0	0	0	0	0	0	0	0	0	0.51
qrис	0.20	0.17	0.21	0.19	0.33	0.26	0.21	0.17	0.12	0
qrисnya	0	0	0	0	0	0	0	0	0	0.25
rencana	0	0	0	0	0	0	0	0.42	0	0
selesai	0.41	0.36	0	0	0	0	0	0	0	0
simple	0	0	0	0.40	0	0	0	0.72	0	0
step	0.49	0	0	0	0	0	0	0	0	0
tarik	0	0	0.43	0	0	0	0	0	0.25	0
tampil	0	0	0.51	0	0	0	0	0	0	0

PEMBOBOTAN KATA DENGAN TF-IDF

Hasil pembobotan dari data penelitian

(3, 614)	0.3069263330717489
(3, 3618)	0.34388417247918895
(3, 3854)	0.34757701852352746
(3, 1248)	0.5839675626732005
(3, 2656)	0.49733736356374353
(3, 256)	0.27994204887568286
:	:
(4186, 2280)	0.2950356305730202
(4186, 1973)	0.2688605437723009
(4186, 3905)	0.241124033516752
(4186, 150)	0.35584989737236844
(4186, 3819)	0.34549886420170267
(4186, 927)	0.432644338088754
(4186, 742)	0.48734225067900927
(4187, 1324)	1.0

Klasifikasi MK-NN

dari 4.191 data setelah pre-processing, data dibagi menjadi 80% data latih yang berjumlah 3.352 komentar dan 20% data uji yang berjumlah 839 komentar.

dari data contoh juga dilakukan pembagian data latih untuk 8 komentar pertama, dan 2 komentar terakhir merupakan data uji

Cosine similarity data contoh

	D1	D2	D3	D4	D5	D6	D7	D8
D1	1	0.442	0.154	0.284	0.597	0.048	0.254	0.033
D2	0.442	1	0.268	0.239	0.504	0.040	0.214	0.028
D3	0.154	0.268	1	0.207	0.258	0.047	0.169	0.033
D4	0.284	0.239	0.207	1	0.475	0.046	0.311	0.317
D5	0.597	0.504	0.258	0.475	1	0.080	0.426	0.056
D6	0.048	0.040	0.047	0.046	0.080	1	0.223	0.178
D7	0.254	0.214	0.169	0.311	0.426	0.223	1	0.156
D8	0.033	0.028	0.033	0.317	0.056	0.178	0.156	1

Cosine similarity data penelitian

Matriks Cosine Similarity:

```
[[1.          0.           0.04196907 ... 0.1924415   0.           0.06489819]
 [0.          1.           0.             ... 0.           0.           0. ]
 [0.04196907 0.           1.             ... 0.08167635  0.           0.0626415 ]
 ...
 [0.1924415  0.           0.08167635 ... 1.           0.           0.1262989 ]
 [0.          0.           0.             ... 0.           1.           0. ]
 [0.06489819 0.           0.0626415  ... 0.1262989  0.           1. ]]
```

Klasifikasi MK-NN

Validitas antar data latih

pada data penelitian digunakan nilai K=1 hingga K=10 dan pada data contoh digunakan nilai K=3

$$validity(D1) = \frac{1}{3}(1 + 1 + 1) = 1$$

$$validity(D2) = \frac{1}{3}(1 + 1 + 1) = 1$$

:

:

:

$$validity(D7) = \frac{1}{3}(1 + 1 + 1) = 1$$

$$validity(D8) = \frac{1}{3}(1 + 1 + 1) = 1$$

$$validity = [1, 1, 1, 1, 1, 1, 1, 1]$$

K	Validity
1	[1, 1, 1, 1, 1, 1, 1, 1, 1, 0, ...]
2	[1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5, ...]
3	[1, 1, 1, 1, 0.6, 1, 1, 1, 1, 0.6, ...]
4	[1, 1, 1, 1, 0.75, 1, 1, 1, 1, 0.75, ...]
5	[1, 1, 1, 1, 0.8, 0.8, 1, 1, 1, 0.8, ...]
6	[1, 1, 1, 1, 0.83, 0.66, 1, 1, 1, 0.66, ...]
7	[1, 1, 1, 1, 0.857, 0.571, 1, 1, 1, 0.714, ...]
8	[1, 1, 1, 1, 0.875, 0.625, 1, 1, 1, 0.75, ...]
9	[0.88, 1, 1, 1, 0.88, 0.55, 1, 1, 1, 0.66, ...]
10	[0.9, 1, 1, 1, 0.9, 0.5, 1, 1, 1, 0.7, ...]

Klasifikasi MK-NN

Cosine distance data latih dengan data uji

pada data contoh

	D9	D10
D1	0.69	1.0
D2	0.56	0.815
D3	0.50	1.0
D4	0.86	1.0
D5	0.49	1.0
D6	0.95	0.453
D7	0.90	1.0
D8	0.97	1.0

pada data penelitian

Matriks Cosine Distance:

```
[[0.94343675 1. 1. ... 1. 0.944364 0.96397327]
 [1. 1. 1. ... 1. 1. 1. ]
 [1. 1. 1. ... 1. 0.98502719 0.9871361 ]
 ...
 [1. 1. 1. ... 1. 0.9623464 1. ]
 [1. 1. 1. ... 1. 1. 1. ]
 [1. 1. 0.88103006 ... 1. 1. 1. ]]
```

Klasifikasi MK-NN

Weight voting sebagai penentu kelas data uji

pada data contoh

	D9	D10
D1	0,680	0,687
D2	0,761	0,960
D3	0,676	0,680
D4	0,820	0,838
D5	0,685	0,695
D6	0,843	1,020
D7	0,709	0,805
D8	0,696	0,790

pada data penelitian

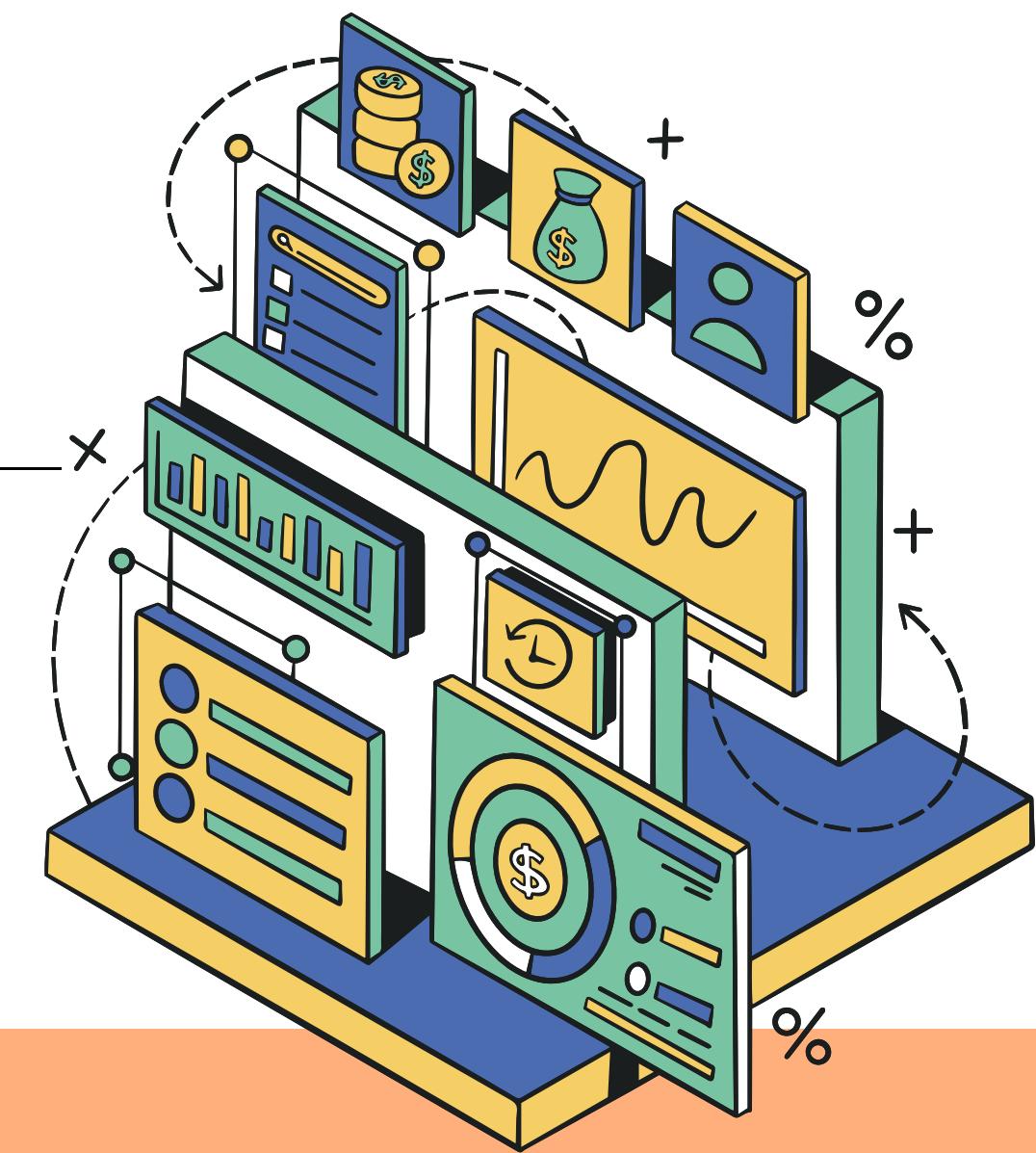
Matriks Weight Voting:

```
[[0.69279101 0.66666667 0.66666667 ... 0.66666667 0.69234625 0.68307258]
 [0.66666667 0.66666667 0.66666667 ... 0.66666667 0.66666667 0.66666667]
 [0.66666667 0.66666667 0.66666667 ... 0.66666667 0.67338834 0.67243341]
 ...
 [0.66666667 0.66666667 0.66666667 ... 0.66666667 0.6838325 0.66666667]
 [0.66666667 0.66666667 0.66666667 ... 0.66666667 0.66666667 0.66666667]
 [0.66666667 0.66666667 0.7240972 ... 0.66666667 0.66666667 0.66666667]]
```

Evaluasi Model

Nilai K	Akurasi	Precision	Recall
1	86,29%	87,63%	94,18%
2	87,24%	87,55%	95,84%
3	86,05%	85,92%	96,34%
4	86,05%	85,60%	96,84%
5	85,45%	84,68%	97,34%
6	85,33%	84,31%	97,67%
7	84,98%	83,80%	98,00%
8	84,38%	83,30%	97,84%
9	84,38%	83,30%	97,84%
10	84,26%	83,00%	98,17%

- Nilai akurasi terbaik terdapat pada nilai K=2 sebesar 87,24%
- Nilai precision terbaik terdapat pada nilai K=1 sebesar 87,63%
- Nilai Recall terbaik terdapat pada nilai K=10 sebesar 98,17%



Hasil Confusion Matrix K=2

Hasil Prediksi	Hasil Aktual	
	Positif	Negatif
Positif	577	82
Negatif	25	155



04

**KESIMPULAN
DAN
SARAN**

KESIMPULAN

1

Analisis Modified K-Nearest Neighbor (MK-NN) dapat digunakan pada klasifikasi penilaian aplikasi wondr by BNI dengan menggunakan beberapa tahapan dari scraping data pada Google Play Store, pre-processing data, pembobotan kata dengan TF-IDF, perhitungan cosine similarity pada data latih, perhitungan validitas pada data latih, perhitungan cosine distance antara data latih dan data uji, dan perhitungan weight voting untuk menentukan label pada data uji.

2

User pada wondr by BNI dari 4.191 data, diperoleh 3.056 data berlabel positif dan 1.135 data berlabel negatif. Kata-kata yang sering dibahas user diperoleh pada wordcloud yang memiliki ukuran besar pada kata yang ditampilkan. Pada sentimen positif kata yang sering dibahas adalah “aplikasi”, “wondr”, “bni”, “mudah”, “bagus”, “transaksi”, “mantap”, dan “keren”. Kata-kata yang sering dibahas user pada sentimen negatif adalah “aplikasi”, “bni”, “masuk”, “gak”, “verifikasi wajah”, dan “transaksi”.

3

Hasil akurasi dari analisis Modified K-Nearest Neighbor (MK-NN) tertinggi diperoleh pada nilai K=2 sebesar 87,24% dan nilai precision 87,55% dan nilai recall sebesar 95,84%. Hasil akurasi yang diperoleh menunjukkan tingkat akurasi yang baik

SARAN

1

Bagi pihak BNI, hasil informasi dari komentar-komentar user yang berbentuk negatif dapat dijadikan bahan evaluasi dalam peningkatan kepuasan pengguna dan memberikan pelayanan semaksimal mungkin, serta untuk pengembangan pembaruan aplikasi wondr by BNI.

2

Bagi peneliti selanjutnya, dapat menggunakan pendekatan machine learning lain sebagai pembanding performa algoritma Modified K-Nearest Neighbor (MK-NN) untuk mengklasifikasikan penilaian wondr by BNI.

Terima Kasih

presented by Arifatul f.e