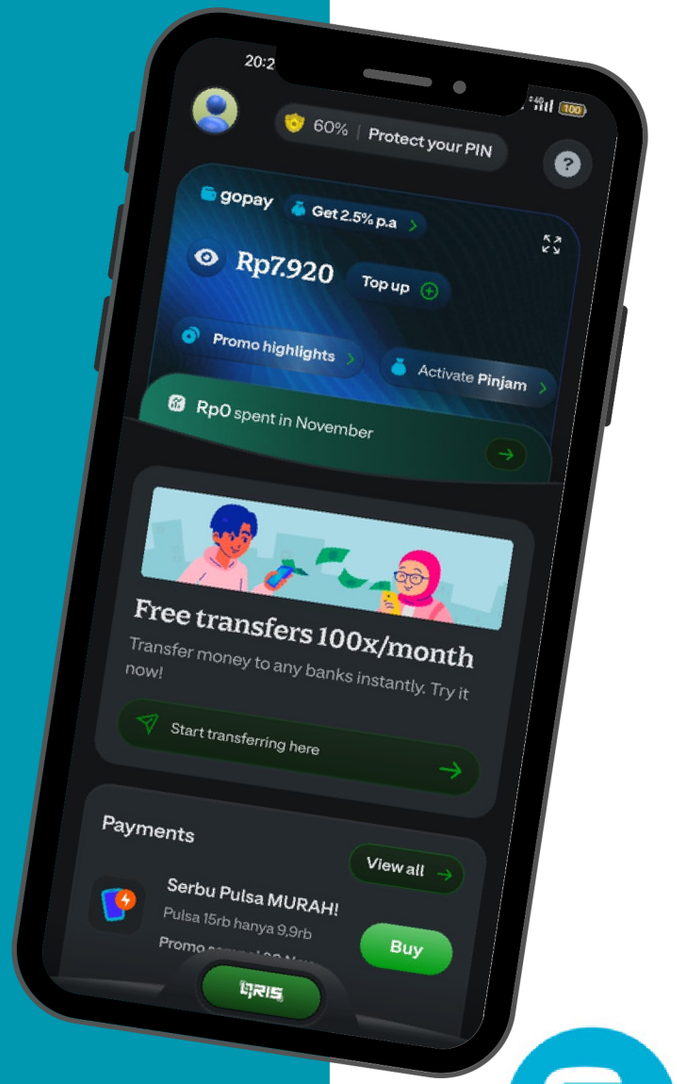
Final Project

Sentiment Analysis of GoPay Application Review from Google Play Store



Presented by:

Aziz Nur Rohman

Data Science Batch 22

ibimbing

gopay

ABOUT ME

MY NAME IS AZIZ NUR ROHMAN

An aspiring data scientist who is in process to switch career into data science field . Currently, learning data science in dibimbing.id bootcamp and expected to be completed in February 2024. Interested in Data Analysis, Machine Learning, and Natural Language Processing (NLP) related.

EDUCATION

- Chemical Engineering, University of Pembangunan Nasional "Veteran" Yogyakarta (2015 -2019)
- Data Science, dibimbing.id (Jul 2023 - Present)

PROFESSIONAL WORK EXPERIENCE

- Quality Control Technician, PT. Pindo Deli Pulp and Paper Mills (May 2021 - Jul 2023)

SKILLS

Python | SQL | Data Cleaning | Data Analysis | Machine Learning | Web Dashboard

INTEREST

Natural Language Processing

aziznurrohman.anr@gmail.com
linkedin.com/in/aziznurrohmananr
github.com/aziznrrhmn

Business Understanding

Background

- GoPay app was introduced in various cities in Indonesia starting from April 2023. Then, It was launched nationally on July, 26 2023.
- It is a lightweight application for transferring to any bank and paying for anything.
- In addition, Providing payment for various needs starting from buying phone credit and data packages, paying bills, shopping online, and much more.

Problem

Since this application is still relatively young, understanding user responses to the application is essential. This helps company track the the application reputation, and step up user satisfaction. Therefore, it is imperative to conduct sentiment analysis through user reviews.

Business Objective

01 Obtain Insight of general rating and sentimen for the application

02 Understand pattern of sentiment each month from launching to present

03 Investigate why reviewers give either positive or negative sentiment in the last three months

04 Create Machine Learning Model to classify sentiment

Data Understanding

Source

Scraped from Google Play Store.

- Only from Indonesia and in Indonesian language (Bahasa)
- Gather data from first review up to November, 22 2023

Data Requirement

Required Column:

- review ID
- at (Date)
- Score (Rating)
- Review Content

Brief Description

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24909 entries, 0 to 24908
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   reviewId  24909 non-null  object
 1   content   24909 non-null  object
 2   score     24909 non-null  int64
 3   at        24909 non-null  object
dtypes: int64(1), object(3)
memory usage: 778.5+ KB
```

Data Preparation

Flow in this step

1 — Data Cleansing

a. Missing Value Handling
b. Duplicated data Handling
c. Result and action:
- No missing value and duplicated data
- Drop 'reviewId' column

2 — Labelling

the score (rating) is labelled into two classification:
- score > 3, the label is positive
- score <= 3, the label is negative

3 — Transformation

Column 'at' is transformed into month. Then, the column is renamed to be 'month'

4 — Text Processing

Conducted:
a. Case Folding
b. Tokenization
c. Replace slang and word elongation
d. Stopword Removal
e. Stemmization
f. Removing irrelevant words manually

5 — Recheck

To check column target (stemmed_text) condition. The column is utilized for further process (Modelling)

N.B : NLP (Natural Language Processing) is applied in Text Processing

Data Preparation

Before

	reviewId	content	score	at
0	62bd418b-27f9-4bdb-8b86-52d42e55a982	Sangat membantu sekali	5	2023-11-21 10:53:03
1	b799bf7a-6eb5-4dc3-8a72-8af417edd1f8	Aplikasinya bagus fitur fiturnya gampang dipak...	5	2023-11-21 10:50:41
2	5f25e5ae-b677-4478-bd95-a25f3fef4d7c	Adanya gopay jadi gampang transaksi apa pun ma...	5	2023-11-21 10:49:32
3	4c6e3e4e-a64b-4a11-b9f6-dd8fd3ea57e6	transaksi lebih mudah dengan aplikasi gopay	5	2023-11-21 10:45:21
4	656dbeac-2246-462a-891b-075682f9fbfb	gak black Ada Vila paling3×	5	2023-11-21 10:44:56
5	b5e159c1-2494-4a8b-aa44-a73eeeaacb47	Bagus banget aplikasinya, nggatau dari dulu be...	5	2023-11-21 10:44:32
6	92799bc1-77fb-449c-84c9-16887e15cba2	Sangat membantu.... Biyaya tf juga sangat mura...	5	2023-11-21 10:42:43
7	91ef23e6-7226-4ebb-a1fe-42de77cc6b98	Aplikasi sangat bangus	5	2023-11-21 10:41:44
8	b985d048-45b4-4522-966c-d0ab9c06eb9f	Aplikasii yang sangat berguna bangett , kalo m...	5	2023-11-21 10:41:19

After

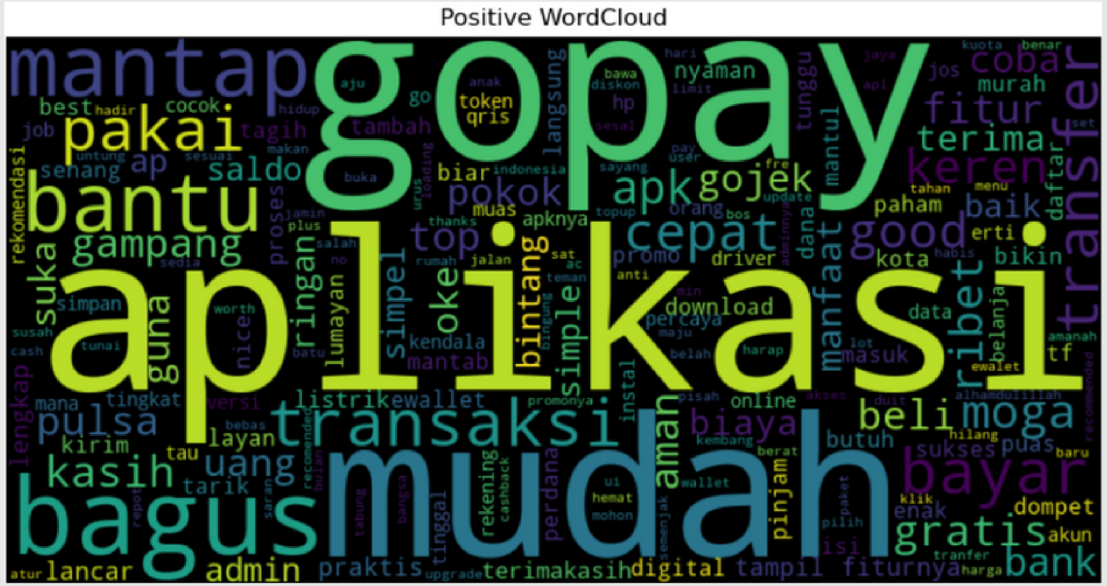
	content	score	label	month	clean_text	token_text	stemmed_text
0	Sangat membantu sekali	5	positive	November	sangat membantu sekali	[membantu]	bantu
1	Aplikasinya bagus fitur fiturnya gampang dipak...	5	positive	November	aplikasinya bagus fitur fiturnya gampang dipak...	[aplikasi, nya, bagus, fitur, fiturnya, gampan...	aplikasi bagus fitur fiturnya gampang pakai ok...
2	Adanya gopay jadi gampang transaksi apa pun ma...	5	positive	November	adanya gopay jadi gampang transaksi apa pun ma...	[nya, gopay, gampang, transaksi, bayar, listri...	gopay gampang transaksi bayar listrik jajan ga...
3	transaksi lebih mudah dengan aplikasi gopay	5	positive	November	transaksi lebih mudah dengan aplikasi gopay	[transaksi, mudah, aplikasi, gopay]	transaksi mudah aplikasi gopay
4	Bagus banget aplikasinya, nggatau dari dulu be...	5	positive	November	bagus banget aplikasinya nggatau dari dulu ber...	[bagus, banget, aplikasi, nya, nggatau, bergun...	bagus aplikasi guna tinggal pakai gopay pakai ...
5	Sangat membantu.... Biyaya tf juga sangat mura...	5	positive	November	sangat membantu biyaya tf juga sangat murah ma...	[membantu, biyaya, tf, murah, mantap]	bantu biyaya tf murah mantap
6	Aplikasi sangat bangus	5	positive	November	aplikasi sangat bangus	[aplikasi, bangus]	aplikasi bangus
7	Aplikasii yang sangat berguna bangett , kalo m...	5	positive	November	aplikasii yang sangat berguna bangett kalo ma...	[aplikasi, berguna, banget, kalo, beli, pulsa,...	aplikasi guna beli pulsa paket data bayar list...
8	Mempermudah dalam transaksi	5	positive	November	mempermudah dalam transaksi	[mempermudah, transaksi]	mudah transaksi
9	apk yang bagus dan sangat mempermudah dalam tr...	5	positive	November	apk yang bagus dan sangat mempermudah dalam tr...	[apk, bagus, mempermudah, transaksi]	apk bagus mudah transaksi

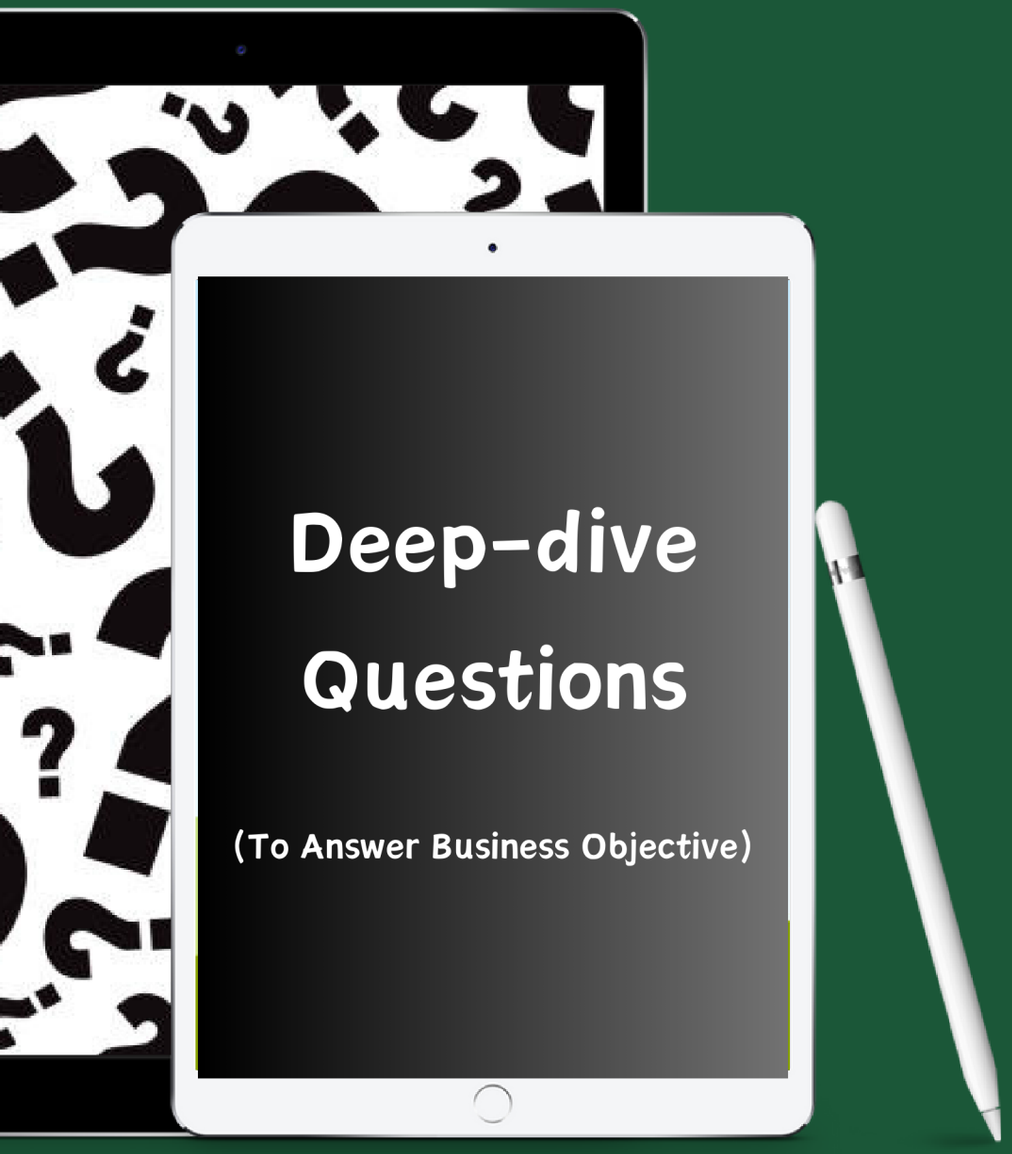
```
<class 'pandas.core.frame.DataFrame'>
Index: 22853 entries, 0 to 24908
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   content       22853 non-null  object
 1   score         22853 non-null  int64
 2   label         22853 non-null  object
 3   month         22853 non-null  object
 4   clean_text    22853 non-null  object
 5   token_text    22853 non-null  object
 6   stemmed_text  22853 non-null  object
dtypes: int64(1), object(6)
memory usage: 1.4+ MB
```

After this step, the dataset is missed 8.25% from original data. However, It is still acceptable

Exploratory Data Analysis (EDA)

Word Cloud


Positive WordCloud
gopay
aplikasi
mudah
bagus
mantap
bantu
pakai
transaksi
transfer
cepat
good
keren
fitur
terima
gojek
apk
pokok
gampang
simple
simpel
ringan
kasih
pulsa
uang
bayar
beli
moga
gratis
admin
praktis
terimakasih
digital
tampil
fiturnya
bank
lancar
manfaat
ribet
biaya
aman
pinjam
download
kendala
mantab
listrikewallet
online
kirim
layan
tarik
tau
kota
driver
promo
kolo
bintang
top
biar
pokok
muas
apknya
saldo
pakai
baik
paham
erti
bikin
tf
dompet
akun
sukses
puas
butuh
masuk
data
percaya
nyaman
murah
cobar
token
qris
langsung
tambah
tunggu
mantul
dana
daftar
menu
tahan
update
indonesia
sayang
user
pay
fre
orang
bos
thanks
urus
loading
jalan
rumah
teman
anti
instal
promonya
akses
berat
kembang
hilang
baru
klik
enak
hemat
mohon
wallet
cashback
rekening
perdana
tinggal
pinjam
isi
upgrade
tranfer
harga
atur
uninstal
respon
tambah
butuh
qris
muncul
foto
tolak
bintang
bagus
upgrade
tunggu
tf
denda
pin
rugi
sesuai
notifikasi
tokopedia
error
cek
baca
pay
limit
jakarta
hadir
gojek
dana
buruk
isi
kirim
bank
pinjam
beli
uang
susah
baik
data
bayar
coba
saldo
pakai
kasih
kota
masuk
daftar
bantu
ribet
plus
apk
potong
sedia
later
pasuk
pulsa
guna
kendala
hilang
ktp
suka
download
jago
mantap
customer
topup
ovo
instal
cepat
hasil
hapus
buka
akun
ribu
gagal
salah
payah
terima
menu
paket
promo
kecewa
tabung
hubung
kode
paylater
notif
mending
biaya
email
admin
selesai
lunas
go
cs
moga
nomor
aju
suruh
tarik
rekening
sayang
kerja
kembali
juta
bikin
pilih
tolol
balas
terang
aktif
lentap
kartu
bunga
nama
jelas
lancar
duit
akses
hari
solusi
layan
gopaylater
aneh
alas
sulit
jelek
jam
mohon
otp
fitur
status
bilang
update
biar
min
kuota
benar
jos
go
anak
hp
diskon
limit
bawa
tambah
buka
sesuai
untung
makan
sesuai
sayang
user
pay
fre
orang
bos
update
dana
mantul
tunggu
langsung
daftar
menu
tahan

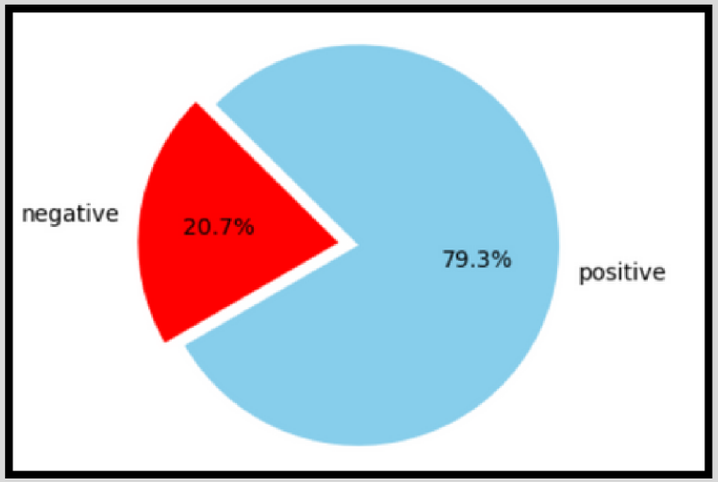


1. How is general rating and sentiment by reviews for GoPay?
2. How pattern of the number of sentiment each month based on reviews?
3. What aspects make users give positive sentiment in the last three months?
4. What aspects make users give negative sentiment in the last three months?

How is general rating and sentiment by reviews for GoPay?

	score
count	22853.000000
mean	4.221940
std	1.435471
min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

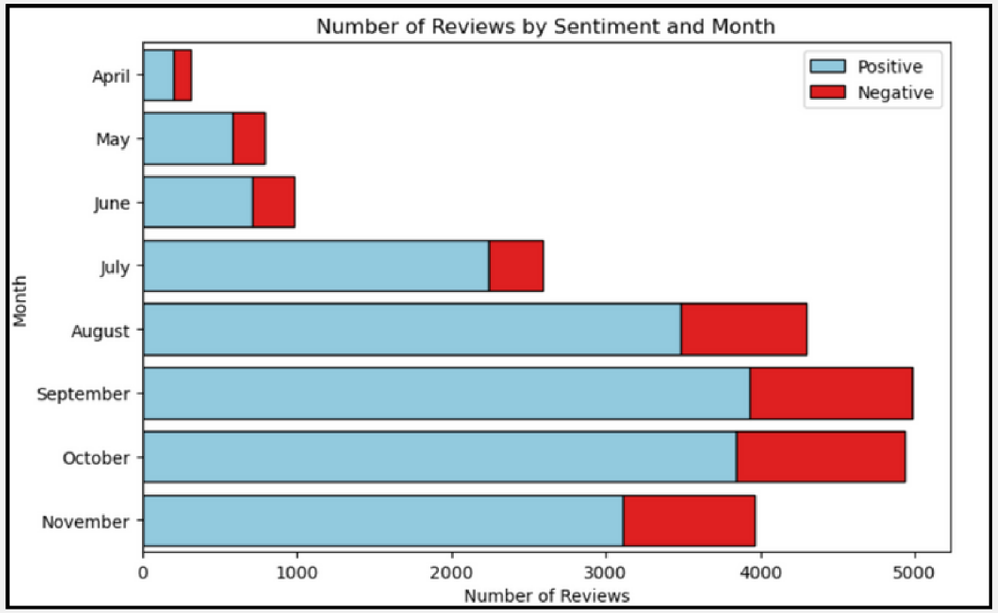
Overall rating for Gopay by review is 4.22. It is superb and give positive sentiment in general

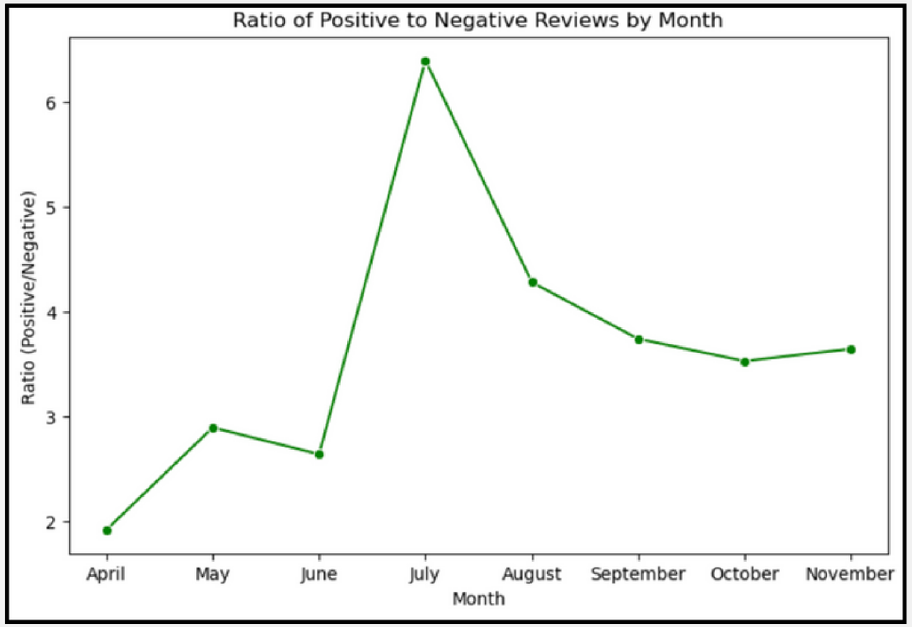


The number of positive sentiment is much greater than negative one. It is 79:21

How pattern of the number of sentiment each month based on reviews?

label	month	negative	positive	ratio
0	April	106	203	1.915094
1	May	202	584	2.891089
2	June	270	712	2.637037
3	July	351	2244	6.393162
4	August	815	3486	4.277301
5	September	1052	3931	3.736692
6	October	1090	3843	3.525688
7	November	854	3110	3.641686

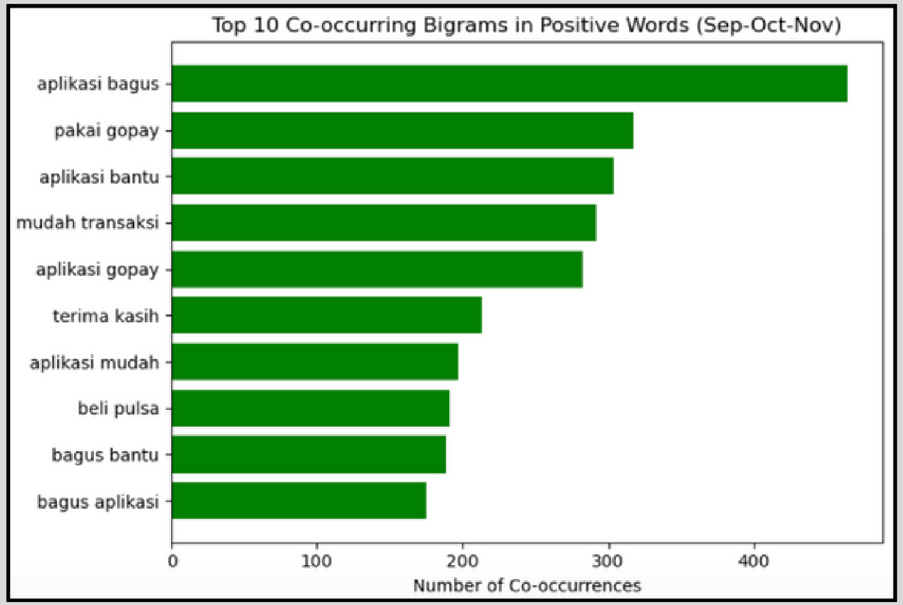


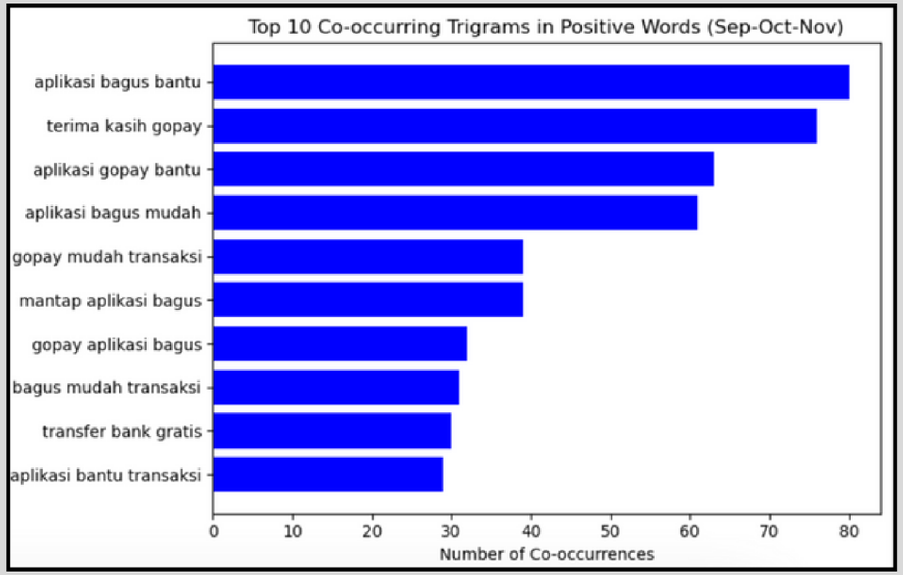


From Observation, It can bee seen:

- In general, Both positive and negative sentiment have uplift the number of reviews every month.
- The greatest number of positive sentimen happens in September, while for negative one, it is in October.
- The ratio in july is the hugest. In this month, the positive sentiment total is 6.3 times negative seniment.

What aspects make users give positive sentiment in the last three months?

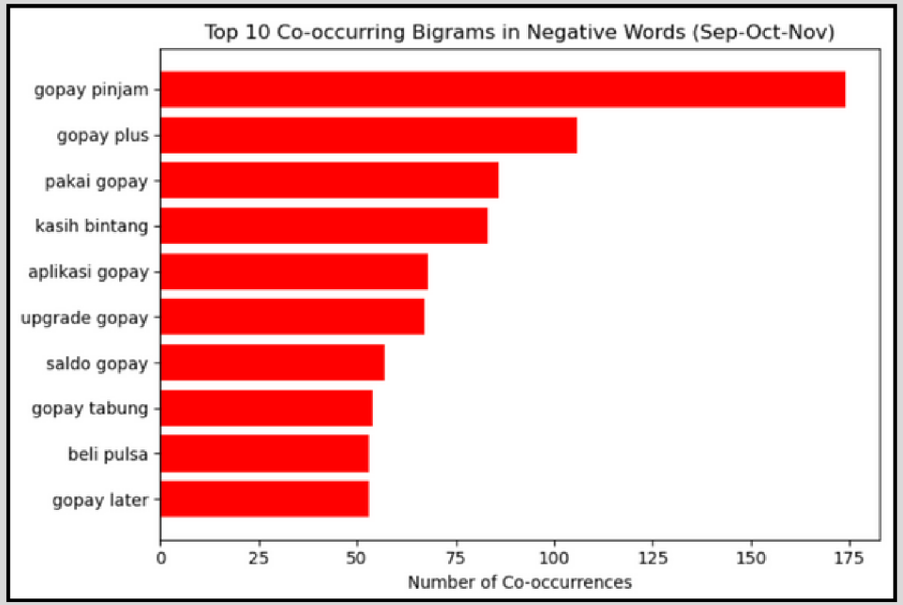


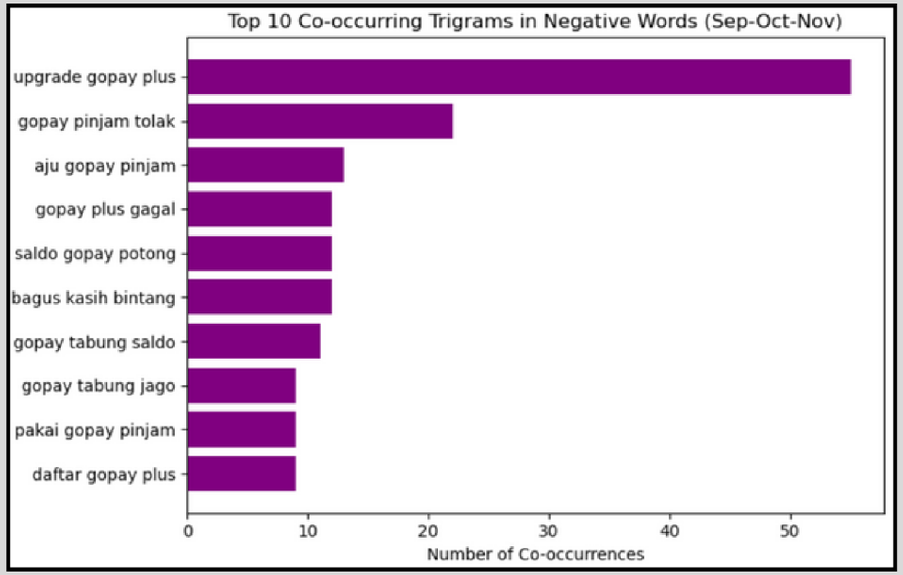


From Bigrams and Trigrams, what the most make users like GoPay is because:

- GoPay is good application which give easiness in transaction.
- Provide free transfer to any bank.

What aspects make users give negative sentiment in the last three months?





According two visual graph, what the most make users dislike GoPay is because:

- There is failure to upgrade GoPay Plus.
- Rejection of requesting GoPay Pinjam.
- The problem of deducted balance.

Modelling and Evaluation

Modelling

Metric

- Expected to have False Positive as small as possible.
- Precision as primary metric. AUC Score (additional metric)
- Threshold:
  Precision >= 90%

Models

Model to compare:

- K-NN
- Logistic Regresion
- Decision Tree
- Random Forest
- Naive Bayes

Methodology

a. Feature: Stemmed_text column,

Target: label column.

Target data is imbalanced

b. Utilizing vectorizer = TF-IDF

c. Comparing:

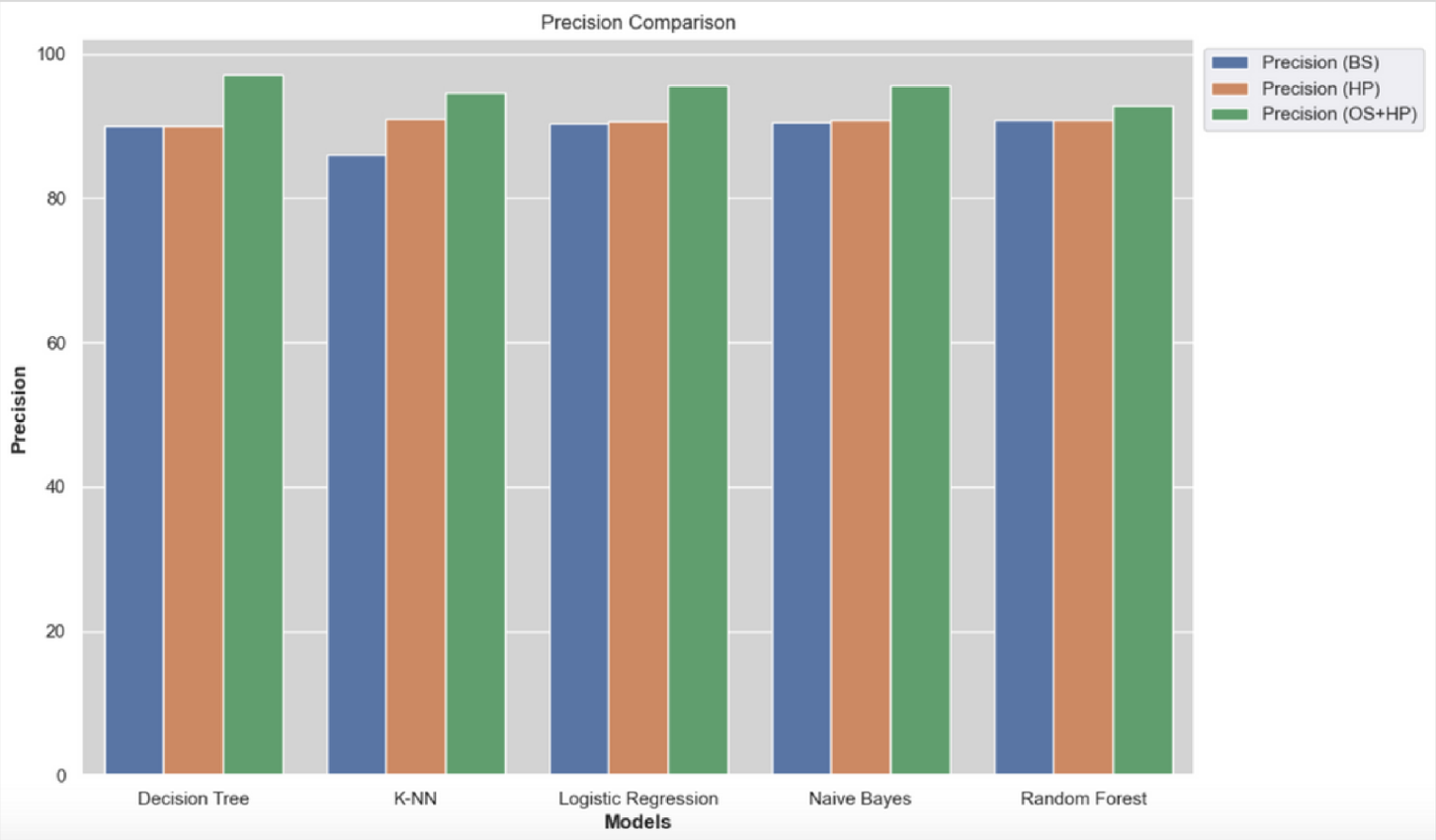
- Baseline (BS)
- CV + HP
- CV + OS +HP

N.B :
CV = Cross Validation, HP = Hyperparameter Tuning, OS = Oversampling

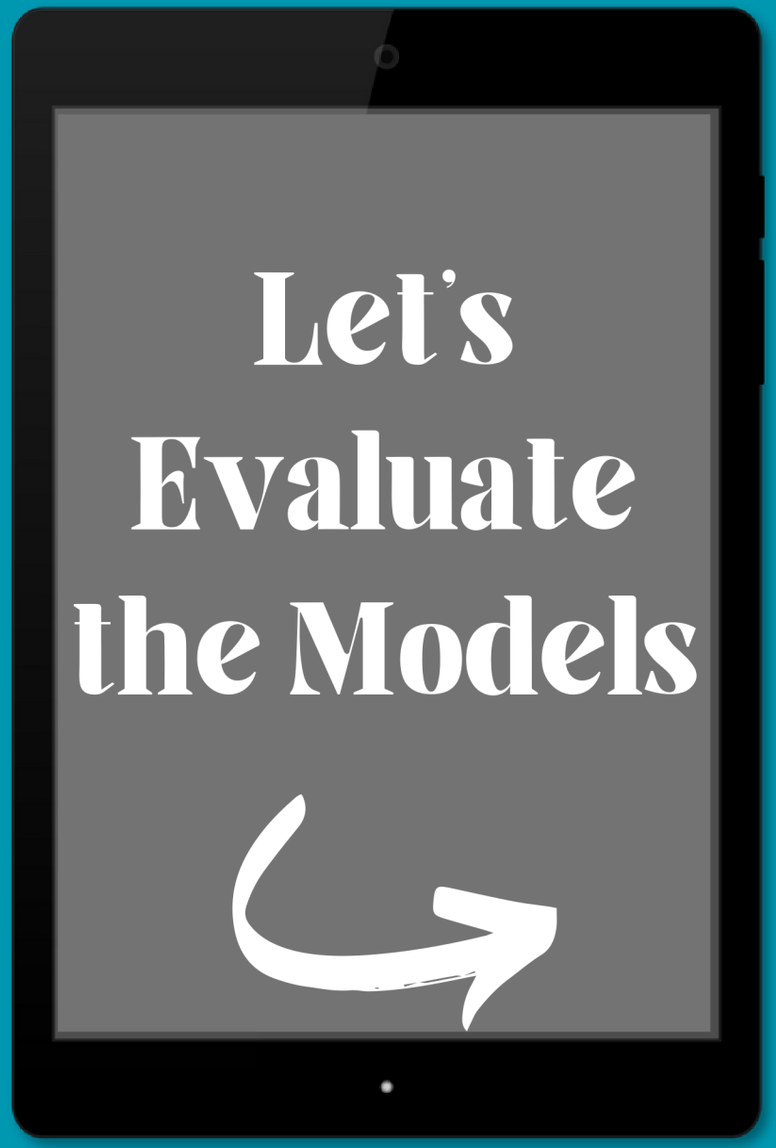
Modelling

Result

Model	Precision (BS)	Precision (HP)	Precision (OS+HP)
Decision Tree	90.029883	89.997282	97.212230
Logistic Regression	90.392208	90.618393	95.667979
Naive Bayes	90.461133	90.926074	95.635041
K-NN	86.069652	90.949033	94.671995
Random Forest	90.846195	90.843758	92.752397



All model have uplift in precision after conducting oversampling + hyperparameter tuning. Hence, it will be evaluated based on the treatment OS + HP



Check Model Condition

Is it Overfitting, Underfitting, or Good fit?

Model	Precision Training (OS+HP)	Precision (OS+HP)	Difference (%)
Decision Tree	98.138453	97.212230	0.943791
Logistic Regression	97.033662	95.667979	1.407432
Naive Bayes	96.715668	95.635041	1.117324
K-NN	98.289161	94.671995	3.680128
Random Forest	98.439397	92.752397	5.777158

Acceptable difference is 10%.

Based on the table:

- All the model is good fit and higher than threshold (>= 90%)
- The top 3 models are Decision Tree, Logistic Regression, and Naive Bayes.

Let's Evaluate the Models

Check Additional Metric

By observing AUC score

Model	Precision (OS+HP)	AUC (OS+HP)
Decision Tree	97.212230	63.498900
Logistic Regression	95.667979	86.773465
Naive Bayes	95.635041	86.680874
K-NN	94.671995	81.518440
Random Forest	92.752397	82.813210

Surprisingly, In spite of having the highest precision, The AUC score of Decision tree is the lowest score. Meanwhile, the other models have good AUC which is more than 80%.

To conclude:

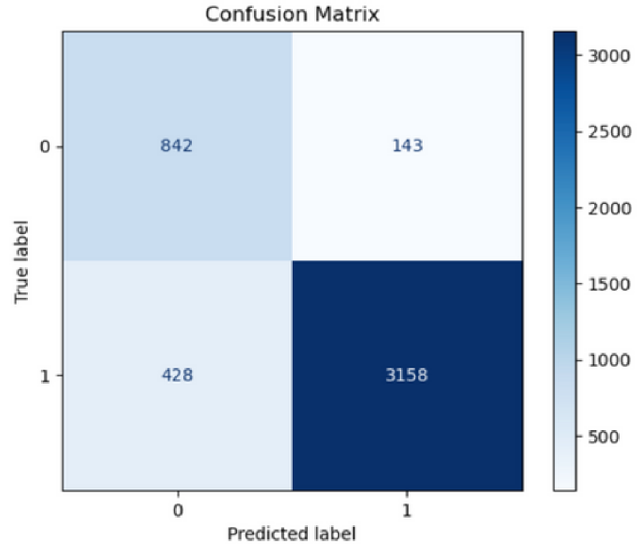
The best Model is Logistic Regression. It has high precision, and the highest AUC score.
Precision : 95.67%, AUC = 86.77%

Let's Evaluate the Models

The best Model

Logistic Regression

```
============================================================
Logistic Regression
Best Estimator:
LogisticRegression(C=1.1999999999999997, random_state=42, solver='liblinear')
============================================================
```

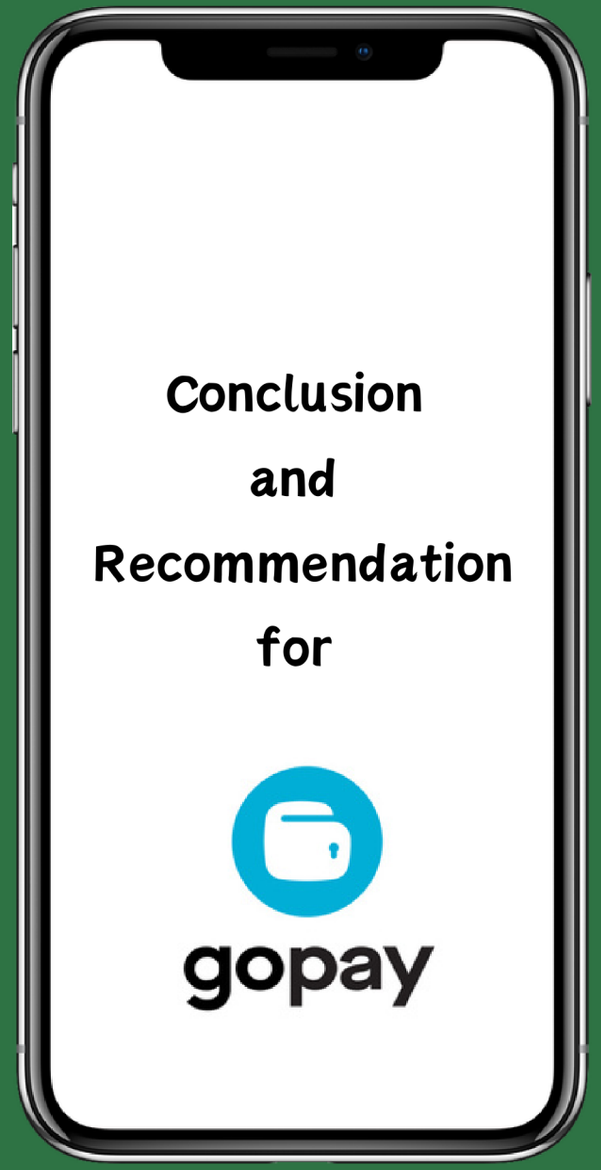


Metric	Value (%)
Precision	95.67
Recall	88.06
F1 Score	91.71
AUC	86.77
Accuracy	87.51

Configuration for Logistic Regression:

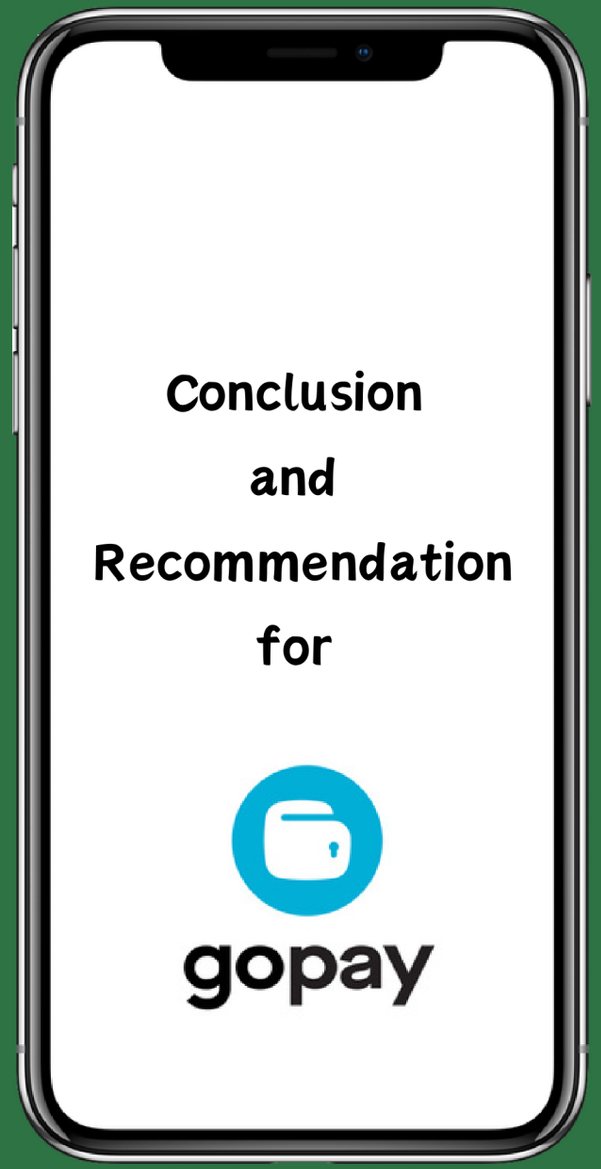
a. Cross Validation (CV)
- Type: Stratified K fold
- n_splits = 5

b. Hyperparameter
- C = 1.2
- Solver = liblinear



Conclusion

1. Overall, General sentiment for GoPay Aplication from launching up to present is good. Based on review, The rating is 4.22. It has 79% positive sentiment and 21% Negative sentiment.
2. The number of positive sentiment is always higher than negative sentiment in each month.
3. Several GoPay users love the application because of easiness in transaction and providing free transfer to any bank.
4. Some users dislike the application due to failure in upgrading Gopay Plus, Rejection of requesting Gopay Pinjam, and the problem of deducted balance.
5. The best machine learning model is Logistic Regression. The value of Precision and AUC respectively is 95.67% and 86.77%



Recommendation

1. Enhance monitoring system to detect potential bug or technical issue related to failure of upgrading Gopay plus and the problem of deducted balance when transaction is not delivered to recipient well.
2. Give specific message clearly why it is failed to upgrade gopay plus to users.
3. Create shortcut information about the cause of gopay pinjam activation is rejected on the menu 'activate pinjam'.
4. Create live chat in working hour on the application to respond user complaint swiftly.
5. Build Chatbot to answer frequently ask of users than only list question and answer on the application.
6. Provide informatian and tips about finance on the application.
7. Create feature such as e-wallet management or e-wallet planning. This can help users divide their money expenditure based on their needs. Thus, this feature can enhance user retention because of versality of the application.

THANK YOU

Let‘s Connect

linkedin.com/in/aziznurrohmananr

github.com/aziznrrhmn