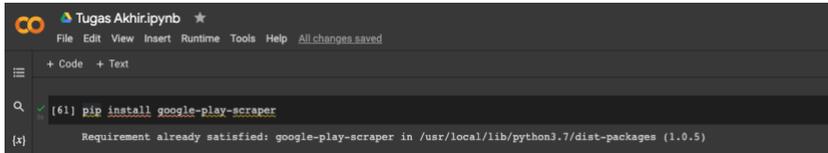


LAMPIRAN

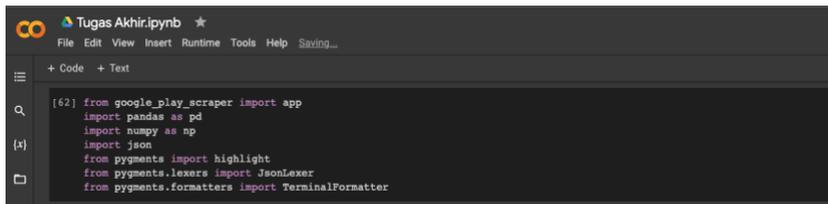
Hal pertama yang dilakukan adalah *download* dan menginstal *Google Play Scraper Package*. Tujuannya untuk mengakses API yang sudah disediakan



```
Tugas Akhir.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[61] pip install google-play-scraper
Requirement already satisfied: google-play-scraper in /usr/local/lib/python3.7/dist-packages (1.0.5)
```

Lampiran 1. *Install Scraper Package*

Dilanjut dengan melakukan proses *import library* dari *pandas*, *numpy*, *json* dan beberapa *library json* lainnya seperti *highlight*, *JsonLaser*, dan *TerminalFormatter*. JSON di sini berguna untuk menampilkan hasil dari ulasan yang diambil ke dalam bentuk JSON. Tujuannya agar lebih mudah dibaca.



```
Tugas Akhir.ipynb
File Edit View Insert Runtime Tools Help Saving...
+ Code + Text
[62] from google_play_scraper import app
import pandas as pd
import numpy as np
import json
from pygments import highlight
from pygments.lexers import JsonLexer
from pygments.formatters import TerminalFormatter
```

Lampiran 2. *Import Library*

Tahapan ketiga yaitu melakukan *scrap* semua ulasan dari aplikasi ternaknesia dengan menggunakan *function reviews_all*. Sehingga akan mengambil semua ulasan tanpa batas. Karena aplikasi ternaknesia itu berasal dari indonesia dan ulasan yang ingin diambil yang berbahasa indonesia, maka untuk parameter *lang* dan *country* diisi dengan *id* untuk *lang*, dan ID untuk *country* dengan kode yang bergantung pada standar ISO 3166 dan ISO 639-1.

Setelah mengambil semua ulasan dari aplikasi, selanjutnya adalah mengubahnya menjadi huruf kecil dengan mengambil isi dari *key content*. Karenan nantinya yang akan difilter adalah *value* dari *content*, dimana *content* adalah komentar yang diberikan oleh pengguna untuk aplikasi di playstore.


```

Tugas Akhir.ipynb
File Edit View Insert Runtime Tools Help @dibungkus

+ Code + Text
[131]: all = df[result]
print("Total Review : " + str(all))

# length = len(str_all)
# print("Total Review yang @Dibungkus : " + str(length))

Total Review : 264

```

Lampiran 5. Jumlah Review Keseluruhan

Mengubah tampilan ulasan dari yang semula JSON ke dalam bentuk tabel dengan mengambil *key* dari *User Name*, *Content*, *Score*, dan *Thumb Up Count* karena ingin mencari tahu siapa yang memberikan komentar, berapa skor yang diberikan dan seberapa banyak pengguna lain yang merasa terbantu atau sependapat dengan ulasan tersebut.

```

Tugas Akhir.ipynb
File Edit View Insert Runtime Tools Help @dibungkus

+ Code + Text
[132]: data = pd.DataFrame()
# data['User Name'] = [a['username'] for a in str_all]
# data['Content'] = [a['content'] for a in str_all]
# data['Score'] = [a['score'] for a in str_all]
# data['Thumb Up Count'] = [a['thumbsUpCount'] for a in str_all]

data['User Name'] = [a['username'] for a in result]
data['Content'] = [a['content'] for a in result]
data['Score'] = [a['score'] for a in result]
data['Thumb Up Count'] = [a['thumbsUpCount'] for a in result]
display(data.head(5))

```

User Name	Content	Score	Thumb Up Count
0	rengas pradana salah satu aplikasi fintech (perbankan) yg rec...	5	9
1	MOCHAMMAD SOLEH suatu aplikasi yg tidak bisa berlaku dioperasya...	1	4
2	Salfu m.sahfuh aplikasiya egak bener min, kalo bisa buat ya...	3	6
3	April Lia dari pertama instal tidak pernah bisa login p...	2	5
4	Drypa Investor bingung di koreksi update yg baru no rekening s...	3	9
...
261	Pengguna Google	bagus	5
262	Bahay Mufiq	bagus	5
263	Muhammad Muzibul Qadri	bagus	5
264	Tu an	mantap	5
265	Bahay Mufiq	bagus	5

264 rows x 4 columns

Lampiran 6. JSON ke Bentuk Tabel

Proses selanjutnya adalah menyimpan tabel dari ulasan tadi ke dalam bentuk file .csv karena selain akan diolah datanya ke dalam bentuk *word cloud* nantinya data tersebut diberikan ke bagian tim data di ternaknesia. File tersebut nantinya tersimpan ke dalam Google Drive dan dihubungkan dengan cara mount ke Google Colab agar file dapat diproses dan dibaca.

```

Tugas Akhir.ipynb
File Edit View Insert Runtime Tools Help @dibungkus

+ Code + Text
[133]: # get reviews of = pd.DataFrame(result)
from google.colab import drive
data_to_csv('reviews.csv', index=False, header=True)
# data_to_csv('reviews.csv', index=False, header=True)

[134]: from google.colab import drive
drive.mount('content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount('/content/drive', force_remount=True).

[135]: read_csv = pd.read_csv('/content/reviews.csv')
label = read_csv['Content']
label[:100]

0 salah satu aplikasi fintech (perbankan) yg rec...
1 suatu aplikasi yg tidak bisa berlaku dioperasya...
2 aplikasiya egak bener min,, kalo bisa buat ya...
3 dari pertama instal tidak pernah bisa login p...
4 bingung di koreksi update yg baru no rekening s...
...
261
262
263
264
265
Name: Content, Length: 264, dtype: object

```

Lampiran 7. Mount Google Drive dan Preview File

Proses mapping selanjutnya adalah melakukan penghapusan karakter, huruf yang berdiri sendiri di awal ataupun diseluruh kalimat, mengganti spasi yang double menjadi satu spasi, dan penghapusan prefix 'b'.

```

import re

processed_features = []

for sentence in range(0, len(labels)):
    # Remove all the special characters
    processed_feature = re.sub(r'[@_!#$%^&*(){}~.;:"]', '', str(labels[sentence]))

    # Remove all single characters
    processed_feature = re.sub(r'([a-z])', '', processed_feature)

    # Remove single characters from the start
    processed_feature = re.sub(r'^([a-z])', '', processed_feature)

    # Substituting multiple spaces with single space
    processed_feature = re.sub(r' +', ' ', processed_feature, flags=re.I)

    # Removing prefixes 'b'
    processed_feature = re.sub(r'b', '', processed_feature)

    # Converting to lowercase
    processed_feature = processed_feature.lower()

    processed_features.append(processed_feature)

```

Lampiran 8. Hapus Karakter

Hasil yang didapatkan dari proses pembersihan kalimat adalah sebagai berikut:

```

[139] labels[300]

0 salah satu aplikasi fintech (investasi) yg recom...
1 aplikasi yg dipertaya,atau salah mana dftar...
2 aplikasi yang akan, kalo report hasil filg...
3 laskal login, rbb daftar, nulll apa ya email ...
4 halng brnaka, updatr yg not rekening kllang.....

Name: Context, Length: 144, drop: object

```

Lampiran 9. Hasil Hapus Karakter

Kalimat dari ulasan yang diberikan nantinya akan dipecah menjadi beberapa kata yang disebut dengan token. Proses ini dapat disebut *Tokenization* yang bisa dilakukan dengan melakukan *import library NLTK* dan mendownload 'punkt'.

```

[140] import nltk
nltk.download('punkt') # Sentence tokenizer
nltk.download('stopwords') # stopwords

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
True

```

Lampiran 10. Download Punkt dan Stopwords

Proses selanjutnya adalah *stopwords* yang dimana hanya mengambil kata yang dianggap penting dan hanya menampilkan lima dataset yang hanya digunakan untuk kalimat yang berbahasa Indonesia.

```

[141] from nltk.corpus import stopwords
stopwords.words('indonesian')
laku
laku
laku
laku

```


Filter awal dilakukan karena masih terlalu banyak kata yang tidak sesuai dengan apa yang dicari, beberapa kata yang perlu dihapus adalah:

- a. Bagus
- b. Ternaknesia
- c. Investasi
- d. Aplikasi
- e. Semoga

```

from google_play_scraper import Sort, reviews_all

# # SCRAPE DATA REVIEW DARI APLIKASI
results = reviews_all(
    com.touristbellea.app,
    skip_pagination=True, # default to 0
    lang='id', # default to 'en'
    sortBy='RATING', # default to SORT_BY_RELEVANT
    filter_score_min=3 # default to None (return all scores)
)

# new_result = json.dumps(results)
# final_result = [{"key": value["text"]} for key, value in results.items()] for mlret in result]
# final_result = pd.DataFrame(final_result)

# # MENYUSUN OUTPUT MENURUT URUTAN RATING
for idx, review in enumerate(results):
    content_text = review["text"]
    content_date = review["date"]
    review_score(review["score"])
    results[idx] = review

# print("\n after : {}".format(results, results[idx]["content"]))

# FILTER BERDASARKAN KATA/KALIMAT YANG CIDER DITAMPILKAN
str_match = [x for x in result if "bagus" not in x["content"] and "ternaknesia" not in x["content"] and "investasi" not in x["content"] and "aplikasi" not in x["content"] and "semoga" not in x["content"]]

# print(str_match)
# print(len(str_match))
    
```

Lampiran 16. Filter Pertama

Filter kedua dilakukan dengan menghilangkan beberapa kata, diantaranya :

- a. Nya
- b. Good
- c. Ok
- d. Ya

```

from google_play_scraper import Sort, reviews_all

# # SCRAPE DATA REVIEW DARI APLIKASI
results = reviews_all(
    com.touristbellea.app,
    skip_pagination=True, # default to 0
    lang='id', # default to 'en'
    sortBy='RATING', # default to SORT_BY_RELEVANT
    filter_score_min=3 # default to None (return all scores)
)

# new_result = json.dumps(results)
# final_result = [{"key": value["text"]} for key, value in results.items()] for mlret in result]
# final_result = pd.DataFrame(final_result)

# # MENYUSUN OUTPUT MENURUT URUTAN RATING
for idx, review in enumerate(results):
    content_text = review["text"]
    content_date = review["date"]
    review_score(review["score"])
    results[idx] = review

# print("\n after : {}".format(results, results[idx]["content"]))

# FILTER BERDASARKAN KATA/KALIMAT YANG CIDER DITAMPILKAN
str_match = [x for x in result if "nya" not in x["content"] and "good" not in x["content"] and "ok" not in x["content"] and "ya" not in x["content"] and "bagus" not in x["content"] and "ternaknesia" not in x["content"] and "investasi" not in x["content"] and "aplikasi" not in x["content"] and "semoga" not in x["content"]]

# print(str_match)
# print(len(str_match))
    
```

Lampiran 17. Filter Kedua

Filter ketiga dilakukan dengan menghilangkan beberapa kata, diantaranya :

- a. Mantap

