İstanbul
Bilgi Üniversitesi

CMPE 251: Data Science and Social Media Analysis, Term Project Report

Bachelor of Engineering

YouTube Comment API

Ahmet Emre Sertdemir (CMPE - ECON) 117200037

Ertuğrul Taparcı (IE - CMPE) 117203052

Sina Berk Golech (CMPE - BIOE) 118200041

Mehmet Yağız Ünsal (CMPE) 117200065

Faculty of Engineering and Natural Sciences

Department of Computer Engineering

9 December, 2020

YouTube Comment API Term Project Report

Shallow Blue

Figure 1: Logo of the project group.

Supervisor: Savaş Yıldırım, PhD

Abstract

YouTube is one of the major social media platforms globally and has big data. With data science and social media analytics, data can be interpreted and processable via tools like APIs that the platforms offered. The big data that YouTube has a lot of categories such as football, music, etc. This report explains which words are most favorable in terms of interest categories. This project uses YouTube API to get the data and propose a data frame to businesses for the usage of commenting about their assets.

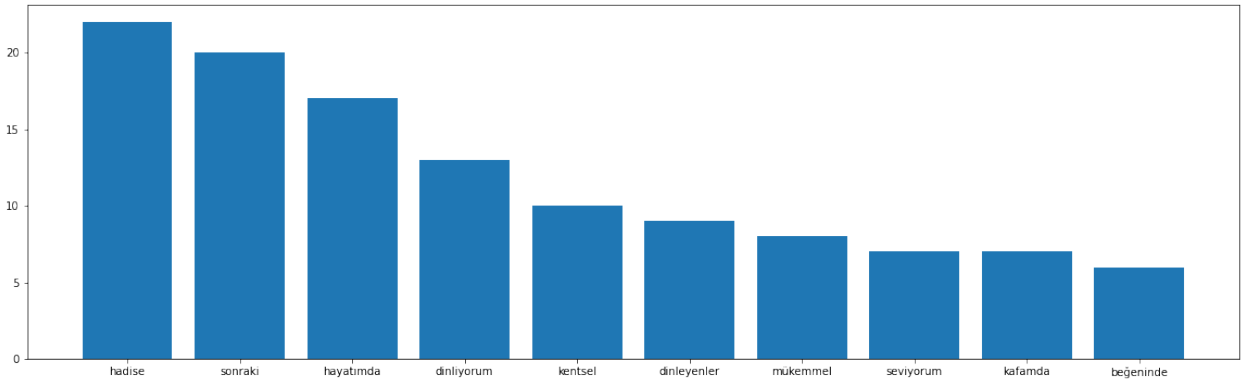

Figure 2: Example of the resulting words.

List of Figures

Contents

Chapter 1

Introduction

This chapter offers the introduction part of the report.

1.1 Introduction

1.1.1 Purpose of the report

This report analyses the data received from YouTube with a specific topic of comment mining with this way the project can offer a recommendation of unique words to use for getting favorited by the users. Additionally, it analyses which topic has a range of favorable words.

1.1.2 Scope of the project

This report provides the company to offer popular words from specific topics. The interests of people get their value from their own topic. So, with this project companies can know the data with the visualisations.

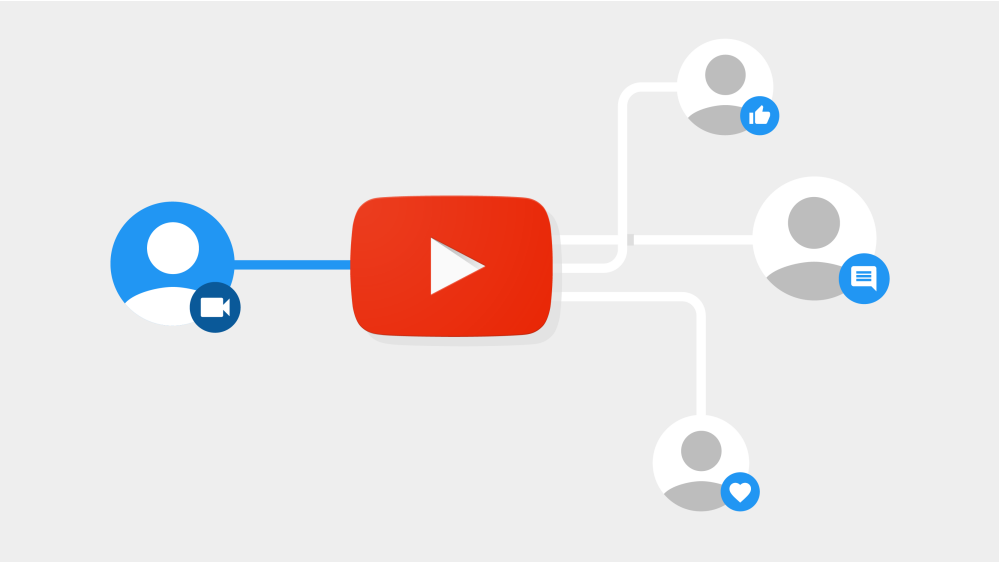

Figure 1.1: Use the API to search for videos matching specific search terms, topics, locations, publication dates, and much more. The APIs search.list method also supports searches for playlists and channels.

(Google. *YouTube API*. `https://developers.google.com/youtube/v3`. Accessed on 2020-12-05. (December 2020))

Chapter 2

Aim of the Project

In this chapter we will look at the aim of the project.

2.1 Aim of the project

This project aims to develop an application which suggest a set of word in a specific domain of subjects. The application analyzes the data driven with the YouTube comments and presents the most favored words.

2.2 Some Data Science Background

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data.

Data science is a "concept to unify statistics, data analysis and their related methods" in order to "understand and analyze actual phenomena" with data.

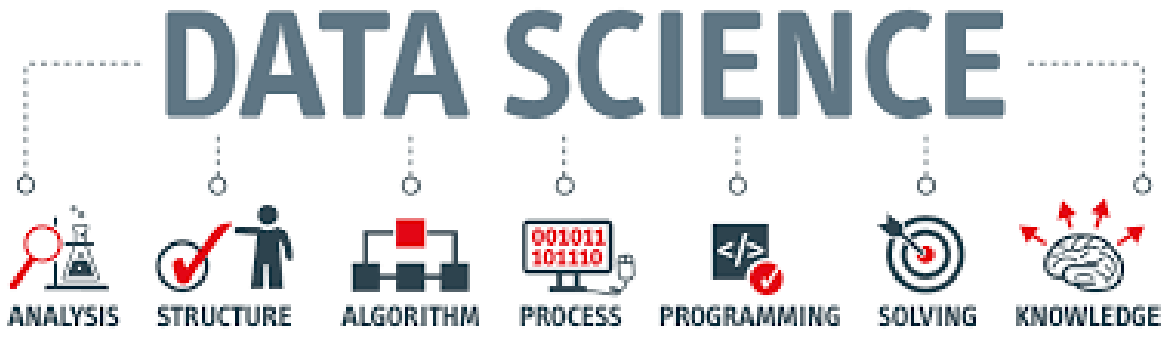

Figure 2.1: Example of the data science applications.

(DAtanami. *Data Science*. `https://www.datanami.com/2018/09/17/improving-your-odds-with-data-science-hiring/`. Accessed on 2020-12-05. (December 2020))

2.3 YouTube API

The YouTube Application Programming Interface allows developers to access video statistics and YouTube channels data via two types of calls, REST and XML-RPC. Google describe the YouTube API Resources as "APIs and Tools that let you bring the YouTube experience to your webpage, application or device.

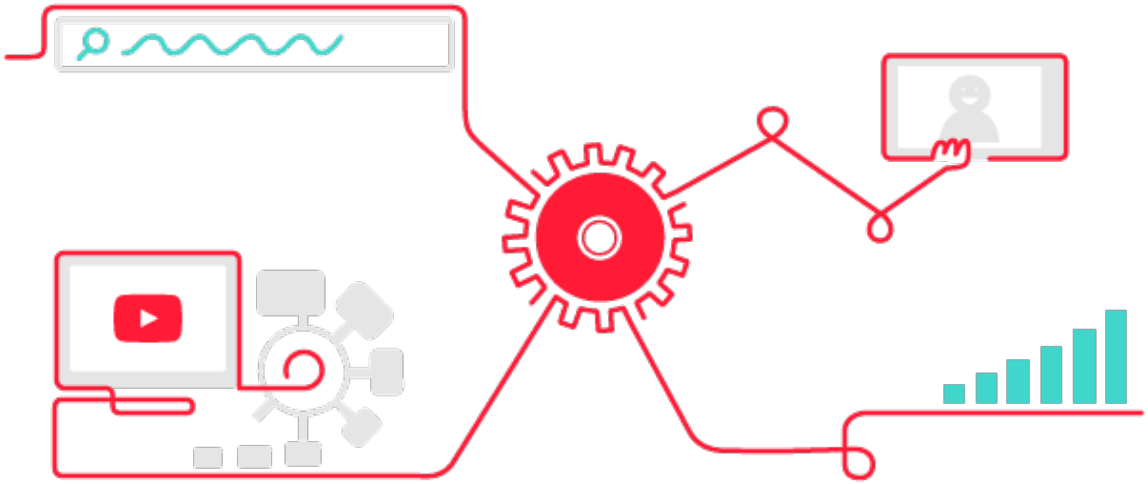

Figure 2.2: Illustrated figure of the YouTube API diagram.

(Google, *YouTube API*, op. cit.)

2.4 GitHub

The project published in the following link $\longrightarrow$ (*GitHubLink*)
(https://github.com/ertugrultaparci/CMPE251/blob/main/YouTube%20Comment-Turkish.ipynb)

Chapter 3

Codes written with explanations

This seems to be the last chapter in this report this chapter we will look at the three experiments with the same method.

3.1 Using API key

3.1.1 Importations

Project uses Google Cloud and import the folowing packages;

- From googleapiclient.discovery "build"
- Pandas
- From youtube_api YouTubeDataAPI
- Nltk
- From collections import Counter
- from nltk.corpus import stopwords

```
from googleapiclient.discovery import build
import pandas as pd
```

```
from youtube_api import YouTubeDataAPI
api_key = "AIzaSyD0hNtrj8OZnhEVvoDFei5j8VP1C4ill6E" #API
youtube = build('youtube', 'v3', developerKey=api_key)
```

Figure 3.1: Figure of the imported libraries and usage of api key.

3.2 Data Collection

Youtube comments are collected by using Youtube API with the help of commentThreads method showed in the figure.

3.3 Data Manupulation

Created DataFrame from dictionary by using "snippet" from the data and reach for the commends.

```
ID = 'DIgbLjkRyko'
```

```
my_data_2 = youtube.commentThreads().list(
      part="snippet",
      videoId= ID,
      maxResults =500,
      textFormat="plainText",
    ).execute()
my_data_2
```

Figure 3.2: Demonstration of the code cell.

```
list_comments = []
for i in song_videos:

    ID = i

    comment_data = youtube.commentThreads().list(
          part="snippet",
          videoId= ID,
          maxResults =100,
          textFormat="plainText",
        ).execute()

    data = pd.DataFrame.from_dict(comment_data['items'])


    for item in data['snippet']:
        c = item['topLevelComment']['snippet']
        list_comments.append(c)

comments = pd.DataFrame(list_comments)
```

Figure 3.3: Usage of one or more videos that have been analyzed.

3.4 Creating dataframe

List has been created and commends added.

```
comments.sort_values(by=['likeCount'],ascending=False)[:3]
```

Figure 3.4: Sort command on the dataframe.

3.5 More than one videos are been analyzed

With this for loop code can extract more than one video. Comments can be taken and ready to interpreted.

```
from collections import Counter
```

```
import nltk
nltk.download("stopwords")
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ertug\AppData\Roaming
[nltk_data]   Package stopwords is already up-
```

```
stp=stopwords.words('turkish')
print(stopwords.words('turkish'))
```

```
['acaba', 'ama', 'aslında', 'az', 'bazı', 'bel
em', 'hep', 'hepsi', 'her', 'hiç', 'için', 'il
, 'sanki', 'şey', 'siz', 'şu', 'tüm', 've', 'v
```

```
onlyComments = df_2[['textDisplay']][:500]
onlyComments
```

Figure 3.5: Demonstration of the NLP unit.

3.5.1 Sorting with respect to like counts

With this code, data can be sorted with respect to like counts. The list is sorted by Like Count from highest like to lowest.

3.5.2 Spliting the text

By using split words, it is the calculated frequency of using it

3.5.3 Cleaning the code

A string has been created named ”all_text”. The comments part of videos have been added to the string and converted to lower case and split. The stopwords have been separated from the whole words.

The words whose length is more than 3 and who don’t include some special characters have been extracted from the word list.

The words whose length is more than 3 and who don’t include some special characters have been extracted from the word list.

3.6 Results of different subjects

This project analyzed three different categories and get the following results;

3.6.1 TV Series

The most popular word in TV series videos is "Dizi". As a conclution this project can suggest to use "Dizi" in the Turkish TV series comments.

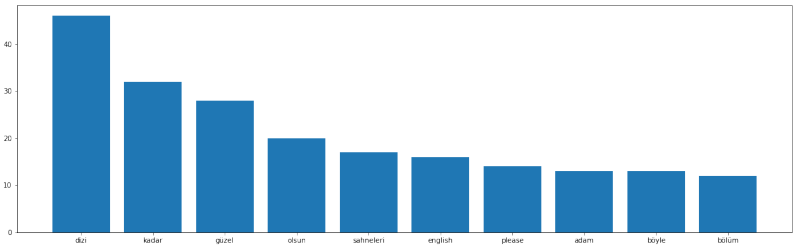

Figure 3.6: Example of the resulting words for TV series.

3.6.2 Music

The most popular word in TV series videos is "Hadise". As a conclution this project can suggest to use "Hadise" in the Turkish TV series comments.

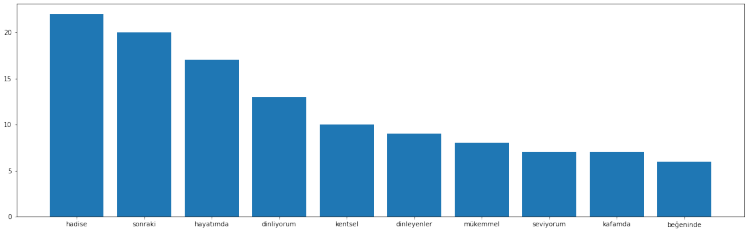

Figure 3.7: Example of the resulting words for music.

3.6.3 Football

The most popular word in TV series videos is "gol". As a conclution this project can suggest to use "gol" in the Turkish TV series comments.

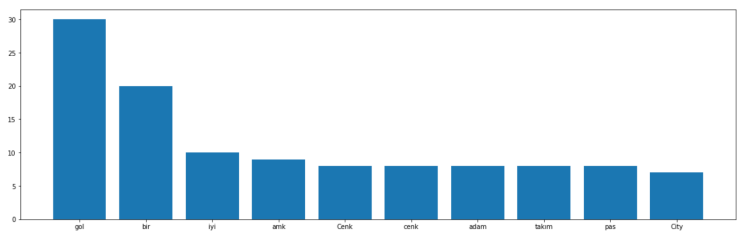

Figure 3.8: Example of the resulting words for football.

Bibliography

DAtanami. *Data Science*. https://www.datanami.com/2018/09/17/improving-your-odds-with-data-science-hiring/. Accessed on 2020-12-05. (December 2020).

Google. *YouTube API*. https://developers.google.com/youtube/v3. Accessed on 2020-12-05. (December 2020).