

Team Members: Kevon C., Payal M., Kevin S.

Movie Genre Prediction

Part 1: Scrapping Movie Data

```
In [ ]: from bs4 import BeautifulSoup
import requests
import os
import pandas as pd
import time
from collections import defaultdict, Counter
```

Getting 50 more links of all movie genres

```
In [ ]: genres = ["Adventure", "Animation", "Biography", "Comedy", "Crime", "Drama", "Family", "Fantasy", "Film-Noir", "His
                "Horror", "Music", "Musical", "Mystery", "Romance", "Sci-Fi", "Sport", "Thriller", "War", "Western"]

url_dict = {}

for genre in genres:
    url = "https://www.imdb.com/search/title/?genres={}&start=51&explore=title_type,genres&ref_=adv_nxt"
    formatted_url = url.format(genre)
    url_dict[genre] = formatted_url

print(url_dict)

{'Adventure': 'https://www.imdb.com/search/title/?genres=Adventure&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Animation': 'https://www.imdb.com/search/title/?genres=Animation&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Biography': 'https://www.imdb.com/search/title/?genres=Biography&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Comedy': 'https://www.imdb.com/search/title/?genres=Comedy&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Crime': 'https://www.imdb.com/search/title/?genres=Crime&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Drama': 'https://www.imdb.com/search/title/?genres=Drama&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Family': 'https://www.imdb.com/search/title/?genres=Family&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Fantasy': 'https://www.imdb.com/search/title/?genres=Fantasy&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Film-Noir': 'https://www.imdb.com/search/title/?genres=Film-Noir&start=51&explore=title_type,genres&ref_=adv_nxt',
 'History': 'https://www.imdb.com/search/title/?genres=History&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Horror': 'https://www.imdb.com/search/title/?genres=Horror&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Music': 'https://www.imdb.com/search/title/?genres=Music&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Musical': 'https://www.imdb.com/search/title/?genres=Musical&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Mystery': 'https://www.imdb.com/search/title/?genres=Mystery&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Romance': 'https://www.imdb.com/search/title/?genres=Romance&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Sci-Fi': 'https://www.imdb.com/search/title/?genres=Sci-Fi&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Sport': 'https://www.imdb.com/search/title/?genres=Sport&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Thriller': 'https://www.imdb.com/search/title/?genres=Thriller&start=51&explore=title_type,genres&ref_=adv_nxt',
 'War': 'https://www.imdb.com/search/title/?genres=War&start=51&explore=title_type,genres&ref_=adv_nxt',
 'Western': 'https://www.imdb.com/search/title/?genres=Western&start=51&explore=title_type,genres&ref_=adv_nxt'}
```