IMDB Score Prediction

Phase 2

TEAM MEMEBER: KIRAN RAM N

NM ID: au723721205024

Innovation :

Before I go for innovation phase let me recall the problem description for clear analysis for the project.

Introduction :

IMDb is the world's most popular and authoritative source for movie, TV, and celebrity content. IMDb users often look at ratings to get an idea of how good movies are, so that they can decide what movies to watch or which ones to prioritize. However, movies that are not yet released don't have ratings, and even the ones with few votes often change as more users vote. We have used linear regressions to predict ratings of movies in general, those predictions rely on features like movie earnings or number of votes, which would not be available for new movies. Instead combined cosine similarities and normalized Euclidean distances with a modified kNN algorithm, which still produced mostly very accurate predictions. This will

provide a way to obtain an estimated rating that's not yet provided by IMDb.

Problem Definition :

- Given a set of movie attributes and features, the task is to predict the IMDb score that a movie is likely to achieve once it is released and rated by users on the IMDb platform. The prediction should ideally be made before the movie is released, as a tool to assist filmmakers, studios, and distributors in making informed decisions about marketing and distribution strategies.

Data source :

The data for analysis is taken from,

https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores

Description of Features in the Dataset:

- Title
- Generes
- Premire
- Runtime

- IMDB score
- Language

Design Thinking:

Empathize:

Understand the problem by empathizing with users and stakeholders.Gather insights into what factors might influence IMDb scores.(e.g., user reviews, movie genre, director's reputation).Identify pain points in current IMDb scoring methods.

Define:

Clearly define the problem statement, e.g., "How might we predict IMDb scores more accurately?"

Set specific goals and success criteria for the prediction model.

Ideate:

Generate ideas for predictive features and models.

Consider using natural language processing (NLP) to analyze user reviews, incorporate movie metadata, and include historical IMDb scores.

Prototype:

Create a prototype of the prediction model, which might involve developing a machine learning algorithm.

Use a dataset containing IMDb scores, movie attributes, and user reviews for training and testing the model.

Test:

Evaluate the model's performance using validation and testing data.

Iterate on the model, refining features and adjusting algorithms as needed to improve accuracy.

Implement:

Integrate the predictive model into IMDb's platform, making it available for users to see predicted scores alongside actual scores.

Scale:

Once the predictive model is proven effective, consider scaling it to include a wide range of movies.

Maintain:

Regularly update the model with new data and adapt to changing user preferences and IMDb's evolving database.

PROGRAM:

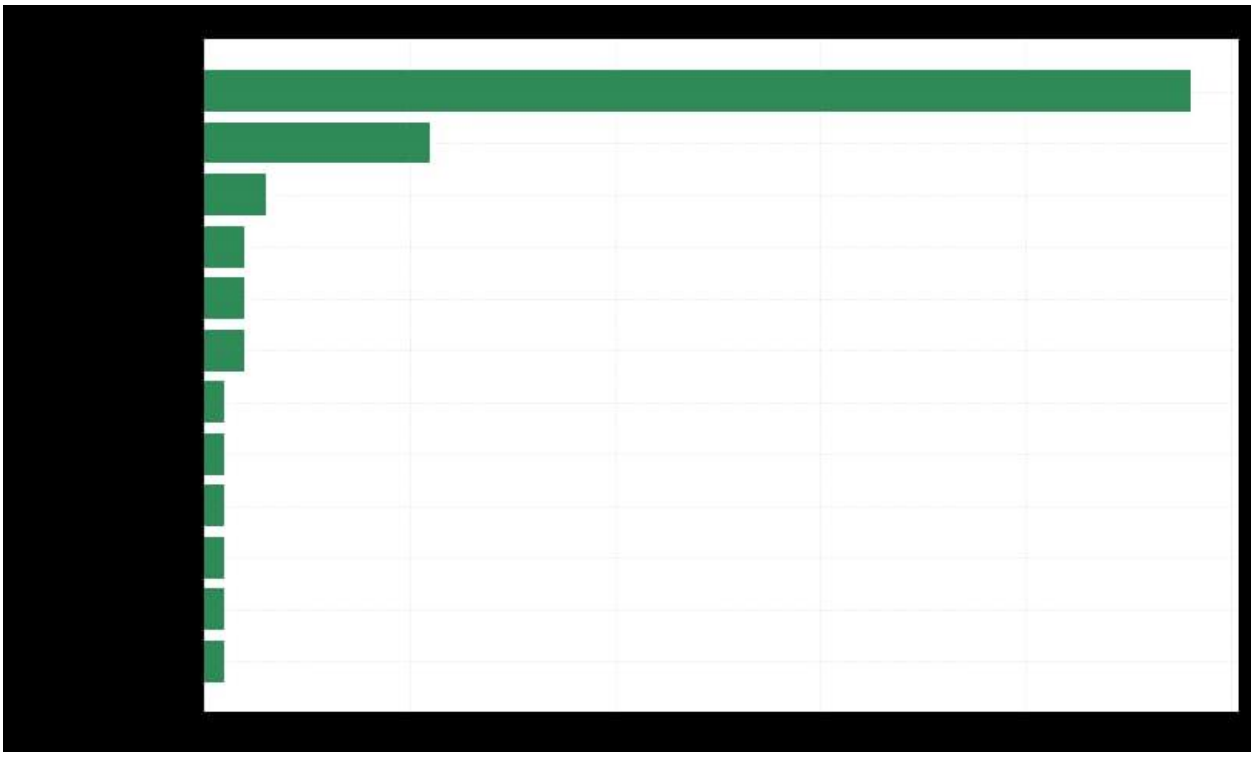
```
Import pandas as pd
Import numpy as np

#Data Visulation
Import seaborn as sns
Import matplotlib.pyplot as plt
Import plotly.express as px
Import plotly.graph_objects as go

From datetime import datetime

Import statsmodels.api as sm

From warnings import filterwarnings
Filterwarnings("ignore")
```

IMDB Score for movies which are lasting more than 2 hours,

```
# Bar Chart
fig, ax = plt.subplots(figsize =(12, 8))

ax.barh(data[data["Runtime"]>=120]["Language"].value_counts().index,

        data[data["Runtime"]>=120]["Language"].value_counts().values,
        color= "seagreen")

for s in ['top', 'bottom', 'left', 'right']:
    ax.spines[s].set_visible(False)

# Remove x, y Ticks
ax.xaxis.set_ticks_position('none')
ax.yaxis.set_ticks_position('none')

# Add padding between axes and labels
ax.xaxis.set_tick_params(pad = 5)
ax.yaxis.set_tick_params(pad = 10)

# Add x, y gridlines
```

```
ax.grid(b = True, color ='grey',
        linestyle ='-.', linewidth = 0.5,
        alpha = 0.2)


# Show top values
ax.invert_yaxis()
# Add Plot Title
ax.set_title('Languages of movies lasting more than 2 hours',
             loc ='center')
plt.show()
```

OUTPUT:



INNOVATION :

- **Deep Learning and Sentiment Analysis:**

❖Implement advanced deep learning techniques and sentiment analysis on user and critic reviews to extract more nuanced insights about movies. This can capture subtle aspects of movies that traditional ratings might miss.

- **Collaborative Filtering:**

- Use collaborative filtering algorithms to recommend movies to users based on their past ratings and preferences. This can enhance user engagement and satisfaction.

- **Contextual Recommendations:**

- Provide IMDb score predictions in the context of individual user preferences. For example, consider a user's favorite genres, directors, or actors when making predictions.

- **Interactive Visualisations:**

- Create interactive data visualizations that allow users to explore IMDb scores in various ways, such as trends over time, genre-specific ratings, or geographic variations in ratings.

- **Crowdsourced Predictions:**

- Involve the IMDb user community in the prediction process by allowing them to make predictions and compare them to actual scores. This can gamify the experience and increase user engagement.

- **Incorporate Social Media Data:**

❖Analyse social media conversations and trends related to movies. Incorporate data from platforms like Twitter, Facebook, and Reddit to gauge public sentiment and buzz around films.

- **AI-Powered Chatbots:**

❖Develop AI chatbots that interact with users to understand their movie preferences and provide IMDb score predictions in a conversational manner.

- **Blockchain for Trust:**

❖Explore blockchain technology to enhance trust and transparency in IMDb scores. Store ratings and reviews on a blockchain for tamper-proof data.

- **Personalized Movie Trailers:**

❖Use AI to generate personalized movie trailers for users based on their preferences. This can pique user interest and influence their movie choices.

- **Predictive Analytics for Filmmakers:**

❖Offer predictive analytics tools for filmmakers and studios to understand how certain factors (e.g., cast, budget, release date) might affect their movie's IMDb score. This could be a premium service.

- **Augmented Reality (AR) Ratings:**

❖Develop AR apps that allow users to view IMDb scores in real-time while watching a movie trailer or

in a physical movie theater. This adds an interactive layer to the movie-watching experience.

- **Voice-Activated IMDb Predictions:**

❖ Enable voice-activated IMDb score predictions through devices like smart speakers and mobile assistants,making it even more convenient for users.

- **Ethical and Inclusive Ratings:**

❖ Consider ethical and diversity aspects when predicting IMDb scores. Develop models that promote fairness and inclusivity in movie ratings.

CONCLUSION :

➢ The world of IMDb score prediction is on the brink of transformation. As technology advances and the film industry evolves, the traditional methods of rating movies are being challenged. The future lies in a more innovative, user-centric, and empathetic approach to predicting IMDb scores.