**Project Report: Ingenious Entertainment advisor**

Innovator: Aryan Verma

Class XI



**------------------------------------------------------------------------------------------------**

**Table of Contents**

1. Introduction
2. Background
3. Experimental Methodology
4. Observations and Future Work
5. Glimpse of the code
6. Acknowledgement and Bibliography
7. Project presentation

**------------------------------------------------------------------------------------------------**

**1. Introduction**

Ingenious Entertainment Advisor is a system that seeks to predict or filter preferences according to the user’s interests. Taking cues from the inputs on movie preferences of the user it applies artificial intelligence, to explore and recommend similar movies enhancing the user experience.

Principle:

This model works on the principle of ***Content-Based Filtering*.** Content-based filtering approaches use a series of discrete characteristics of an item in order to recommend additional items with similar properties. Content-based filtering methods are based specifically on item descriptions and seeks to match the profile of the user’s preferences. It recommends items based on historical data and interests.

**------------------------------------------------------------------------------------------------**

**2. Background**

Human life is fast evolving, with multiple avenues for development and the growth. Our busy lives have reduced time for leisure and entertainment. Movies are an easy medium to take your mind off troubles and relax. However, people are often confused about what they ought to watch.

The Current Outlook-:

More often than not, it is believed that viewers will be likely to choose movies with high ratings and we assume they will pick those.

Fact-:There is however no conclusive evidence that ratings will determine whether more viewers will choose to watch a movie.

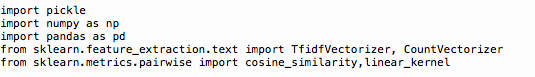
Reason-: This is because people hink differently, have different interests and therefore like different actors, storyline and genres. A rating does not alone inspire a viewer to make a selection. Their moods and tastes also play a role.

Observation-: Often people just remember names of a few of their favourite movies and can’t make a selection. They are happy to be spoilt for choice and promoted by an intelligent source for guidance in line with their tastes.

Conclusion-: Here’s when , Ingenious Entertainment Advisor steps in to solve this problem by using it’s artificial intelligence to recommend the best movie for us by considering some specifications given by a viewer.

**------------------------------------------------------------------------------------------------3. Experimental Methodology**

Step 1: Importing the Python Installed Packages(pip) such as:

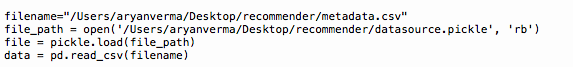


1. Pickle: Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk.

* Then we can store the dataset in a variable (‘for example: file\_path’)
* Then we’ll save the dataset in pickle format in a variable(file)

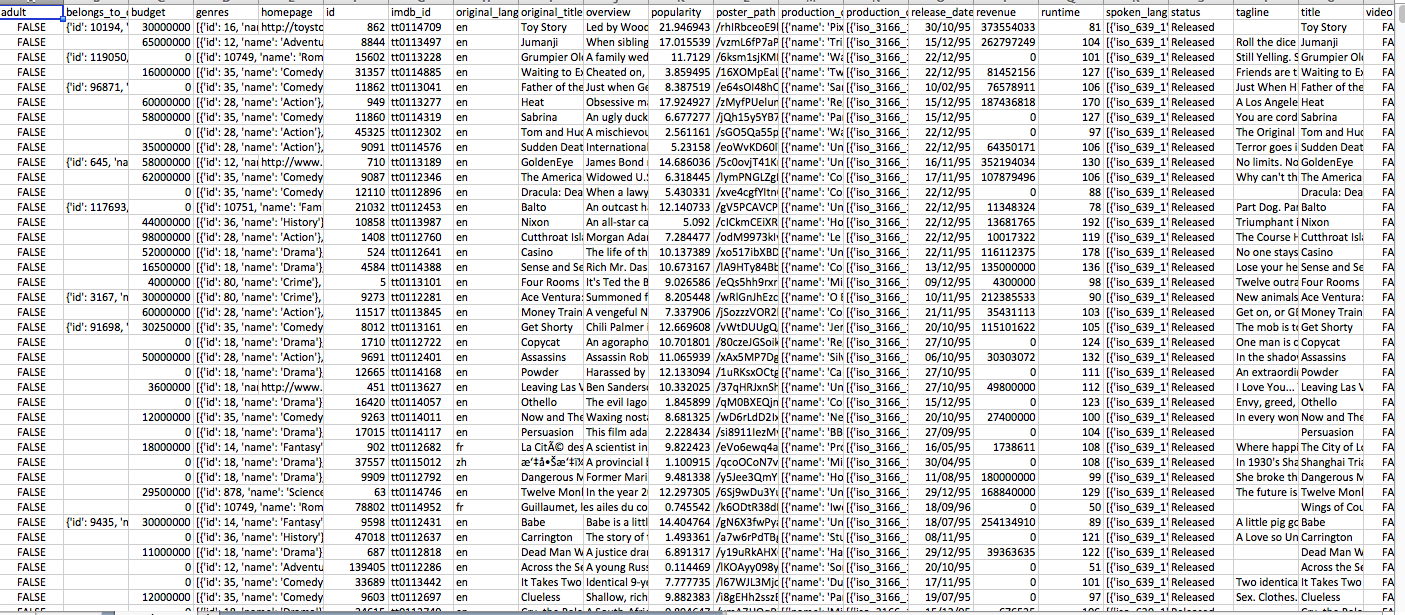
1. Numpy: NumPy is an array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays
2. Pandas: Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
3. Tfid Vectorizer: The TfidfVectorizer will tokenize documents, learn the vocabulary and inverse document frequency weightings, and allow you to encode new documents
4. Count Vectorizer: The CountVectorizer provides a simple way to both tokenize a collection of text documents and build a vocabulary of known words. It also enables you to encode new documents using that vocabulary.
5. Cosine Similarity:Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.
6. Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line.

Step 2: Loading the dataset as follows and converting it into pandas format for further operations:



Step 3: Pre-Processing data

Here is a snapshot of the database we used :



As visible ,there are multiple attributes to each movie such as runtime, star-rating, tagline, story, genere etc…

By pre-processing we need to identify the attributes which would give the most accurate results.

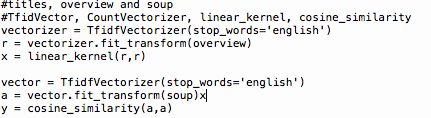
After trying out numerous combinations, overview and soup gave the most accurate results and therefore were used as the basis for checking similarity in the feature comparisons. 

Step 4: Convert the features into vector format so that program could process it using TfidVectorizer or CountVectorizer



Note: The stop\_words attribute can get large and increase the model size when pickling.

Step 5: Finding the similarity between the features of diff films using cosine similarity or linear kernel :



Step 6: Convert the outcomes(titles) in dataframe/series format using panda Then you interchange the position of titles and index

Reason-: This is done because of the simple reason that it is easier to process numericals rather than text

For Example:The movie “Dark Knight Rises” .It is stored with the key 1, where value: Dark

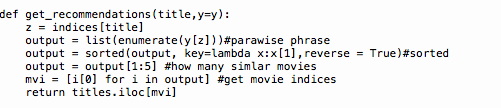
1:Dark Knight Rises

We interchange the key with value

Dark Knight Rises:1 (The key now is Dark Knight Rises and the Title/value is 1 which is easier to process.)

This is done as follows:

Step 7: Create a user defined function to store the process of sorting and para wise phrase

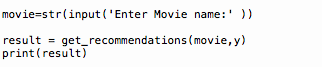


* Under this function execute para-wise phrasing as done below
* Then sort the similarity in descending order to find most similar movies-:
* Then choose how many outcomes(similar movies ) you want-:

Then get movie indices by-:Return movie titles

Note: Pandas provide a unique method to retrieve rows from a Data frame. Dataframe.iloc[] method is used when the index label of a data frame is something other than numeric series of 0, 1, 2, 3....n or in case the user doesn’t know the index label.

Step 8: Ask the user to enter his favourite movie and Execute the function you have created and store it in a variable and Print the variable



**------------------------------------------------------------------------------------------------**

**4.Observations and Future Work**

DATASET-:

Observation -: The dataset, presently available is a little outdated i.e. it does not have the latest movies.

Future Work-: I would like to keep updating the database with the latest movies or in future link it to the official sites of various countries and enable it with automated updating features.

Observation1 -: This function at the moment does not understand the unique abbreviations which users give to movies [for example, Dilwale Dulhania Le Jaenge= DDLJ]

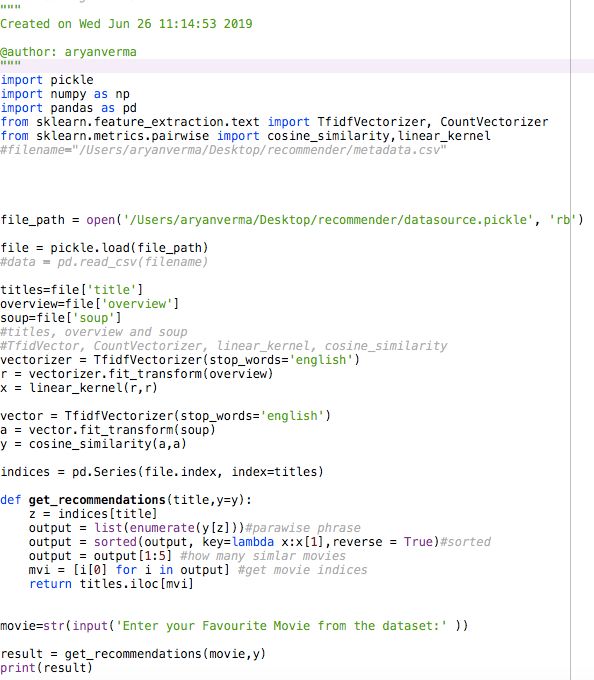
Future Work-: Enhancing the advisor in such a way that it learns from the users diction smartly using Artificial Intelligence to interpret the implications of misspelt or similar words and recommend intuitively acting on user expectations.<< can we put it this way – instead of saying continuously changing vocabulary>>

Future Work-: Providing the project with enhancement functions such as voice recognition and increasing its comprehension and inclusion of multiple languages

Long term goal: The long term goal is to keep evolving through learning and innovation and enrich myself with multidimensional expertise to initiate my own start-up AV tech which creates dreams digitally.

**------------------------------------------------------------------------------------------------**

**5.Glimpse of the Code**



**6.Acknowledgement and Bibliography**

Acknowledgement

I express my special thanks to my diligent and knowledgeable teacher, Dr.Sarvjeet Herald (Phd Computer Science ,Marie Curie University, United Kingdom). With two patents to his name, he gave me direction and always encouraged my innovative mindset promoting intense work effort and patience.

I am also grateful to the founder of Robogenius Academy (Shabnam ma’am), 2018’s India’s best STEM Organisation who gave me the golden opportunity to do this wonderful project on the topic “Ingenious Entertainment Advisor” which also helped me in doing extensive research on areas within Artificial Intelligence.

With their persistent efforts and exceptional mentoring I was able to gain expertise in Python basics and fundamentals, machine learning , decision making capacity of machine(decision tree classifier), work with different python packages and techniques such keras, pandas, tenserflow to an extent that I could comprehend how mighty softwares such Siri and Alexa work.

Their intellectual rigor and research approach will stay with me forever and guide my further studies.

Bibliography:

Software: Python3

Platform : Annaconda

Launcher : Spyder

Sites referred :

1. https://scikitlearn.org/stable/modules/generated/sklearn.feature\_extraction.text.Tf idfVectorizer.html
2. https://www.geeksforgeeks.org/python-extracting-rows-using-pandas-iloc/
3. https://docs.google.com/document/d/1Jy6xrFT3K9OboFKRdseLF8LPMYHnKps0l CeD3VQNPYQ/edit#
4. https://www.designevo.com/apps/logo/?name=yellow-magnifying-glass
5. geeks for geeks.org
6. towardsdatascience.com
7. machinelearningmastery.com

**------------------------------------------------------------------------------------------------**

**<< What is this ??>>**

**Project Representation:**