The topic that we have chosen is Movie Recommendation System with Machine Learning. We are group number 10. I am Pankti Nanavati Roll no-1814045 and my group members are Akhil roll no 1814042 and Jaykumar Panchal roll no 1814044.

2nd Slide

Problem Statement

Movie recommendation syatems are software program that make recommendations to users based on a variety of parameters. These systems forecast the most likely product that users will buy and that they will be interested in. In today's society, online recommendations systems are practically prevalent. Almost all consumer facing websites that we use and visit from Amazon, flipkart to Netflix, offer some sort of recommendation on the content. Example, Flipkart and amazon offer recommendation to user on any product like books to the users based on their previous choices. Facebook offers recommendations of friends to connect with and news websites offer the most relevant article recommendation based on users preferences. These recommendation systems provide huge benefits to salespeople with increased sales and money generation. On the other hand, users who might be the purchasers as well are frequently overwhelmed by a plethora of choices and possibilities in online business experiences, while also having limited resources and time to devote to the choosing process benefit significantly.

In reality, developing a prediction system for a movie database is extremely closely related to developing a recommendation system. The movies that a user is likely to see will be the recommended products in this case. We aim to develop a prediction system that predicts the movie preferences for an unknown user based on some factors. As described in the problem statement, we essentially aim to predict whether a user is likely to watch a movie or not.

3rd Slide

Now, moving on to demonstration of the code. This is the github link which contains the code.

We first import sys which provides various functions and variables that are used to manipulate different parts of the Python runtime environment. sys.path is a built-in variable within the sys module. It contains a list of directories that the interpreter will search in for the required module. Here, it will look for the movierecommendation drive link.

Then we, import various libraries that are required-

Pandas is a **Python package providing fast, flexible, and expressive data structures** designed to make working with "relational" or "labeled" data both easy and intuitive.

Matplotlib is a **cross-platform, data visualization and graphical plotting library for Python** and its numerical extension NumPy.

NumPy arrays **facilitate advanced mathematical and other types of operations on large numbers of data**.

Python's SciPy gives tools for creating sparse matrices **using multiple data structures**, as well as tools for converting a dense matrix to a sparse matrix.

mpl_toolkits. axes_grid1 is **a collection of helper classes to ease displaying (multiple) images with matplotlib**.

Itertool is a module that **provides various functions that work on iterators to produce complex iterators**

Scikit-Learn, or sklearn , is **a machine learning library for Python** that has a K-Means algorithm implementation that can be used instead of creating one from scratch.

The sklearn. metrics module **implements several loss, score, and utility functions to measure classification performance**.

Drive.mount is used to read data form google drive

Then, we import the movies csv file from the drive and display the head of the dataset as seen.

The movies dataset consists of the attribute movieid, title i.e the name of the movie and the genre of that movie.

We import the second csv file from the drive named ratings. The ratings dataset consists of the attributes userid i.e the id of the user who has rate the mobvie, the movieid of the movie, rating of the movie i.e on the scale of 1-5 and the timestamp.

Now, since we are aware of the dataset, we find out the total number of ratings i.e approximately 1Lakh and the total number of movies are 9125.

The further part will be explained by Jaykumar Panchal.