**Introduction to Recommendation System:**

Most of the top companies like Amazon, Netflix, Youtube rely on a recommendation system to serve their customers and anticipate their needs based on their interests. So there is a large need to make a recommendation system that personalizes the product recommendation based on user interest. It not only saves time and effort from the customer side but also makes huge profits for companies.

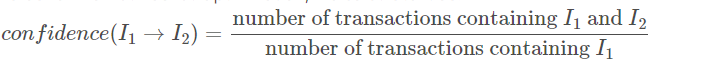
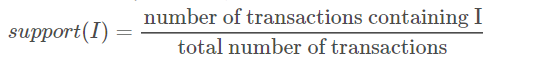
For instance, If some users have watched a movie like Batman and rated it to five stars, it is pretty evident that this user may watch captain America and jungle king movies in the future so It would be better if the user gets the recommendation of a movie depending on the previous movie user has rated.

**Association Rule Mining Concepts :  
Support (Formula):**

Support is the basic and important thing in association rule generating. It tells us the number of transactions in which a particular item is appearing.  
Support of Item I can be calculated by the following formula.

**Confidence:**

Confidence tells us about the likelihood of happening I2 given that I1 has already happened, represented as I1->I2.  
Confidence(I1->I2) can be calculated by the following formula.



**Lift:**

Lift(I1->I2) is also known as gain in the classification model and it is a ratio of the confidence of the rule and the support of the rule(I2).  
Lift(I1->I2) can be calculated by the following formulas :

**Frequent and Non Frequent Itemsets:**

A frequent itemset is a set of items whose Support is more than the threshold value set by the algorithm and a non-frequent itemset is a set of items whose support is less than the threshold value.

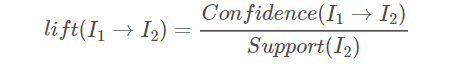
**Apriori Algorithm:**

Apriori algorithm is the classical Association rule mining algorithm for generating rules basically It contains the following steps:

Step:1: set threshold value for support and confidence  
Step:2: select all itemset having support higher than the threshold

Step:3: select all the rules from the above set that have confidence higher than the threshold.

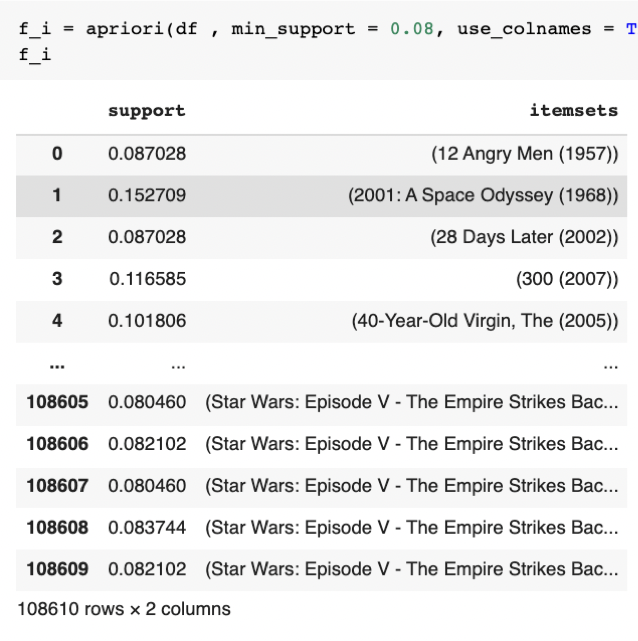
Step:4: now sort all rules in descending order of confidence or lift.



**Approach:**

* ●  We have made a recommendation system for predicting the movies using the apriori algorithm of the association data mining technique.
* ●  To make recommendations we have created each user profile with their respective rating for each movie.
* ●  We have divided each user profile with high-rated movies (i.e. user has given rating >= 3 ) and low-rated movies(i.e. User has given a rating < 3).
* ●  For all high-rated movies for a user, we have made a list of lists and passed it as input to the transaction encoder.
* ●  Transaction encoder will encode into a data frame where each Transaction Ti will represent the user’s most liked movies and then using mlxtend library we will create a list of frequent itemset with a support value of 0.08.





* ●  Now using the apriori function from mlxtend library we will generate a set of rules for all high-rated movies from each user. (Association Rules Mining)
* ●  From the generated set of rules, we will first sort the rules according to the strength of the rule which is majorly decided by lift metric.



* ●  We will next match the antecedents from the rules generated by apriori and then recommend movies based on the consequents generated by association rules.
* ●  Our system will also work on multiple movies for one user.
* ●  In order to make our system more efficient and robust, we have generated 2 M+ rules, and also if none of these rules match it will search for the most favorite genre of the user and will predict the most highly rated movies from a particular genre of the user’s choice.

**Input File Processing :**

* ●  To validate our recommendation system we have tested our recommendation system against 21 users with different likes and dislikes and different tastes in movies.
* ●  Input contains a set of movies watched by a particular user.
* ●  On the basis of the input recommendation system check the input against all the

possible rules for recommendation of movies.

* ●  If in case no rule is found then the recommendation system will predict top rated

movies on the basis of the genre that a particular user is interested in.

* ●  After predicting the movies it will store back the recommendation in a dataframe

which later will be converted into csv format which can be seen in the below snapshot.



**Test Set**

**Precision :**

We are calculating the precision based on the recommendation made by the recommendation system. We have compared the genres of the recommended movies with the input list of movies. The total precision will be the average for all the users.

For our set of 21 users the average precision comes out to be 94 percent which can be seen in the snapshot below.

