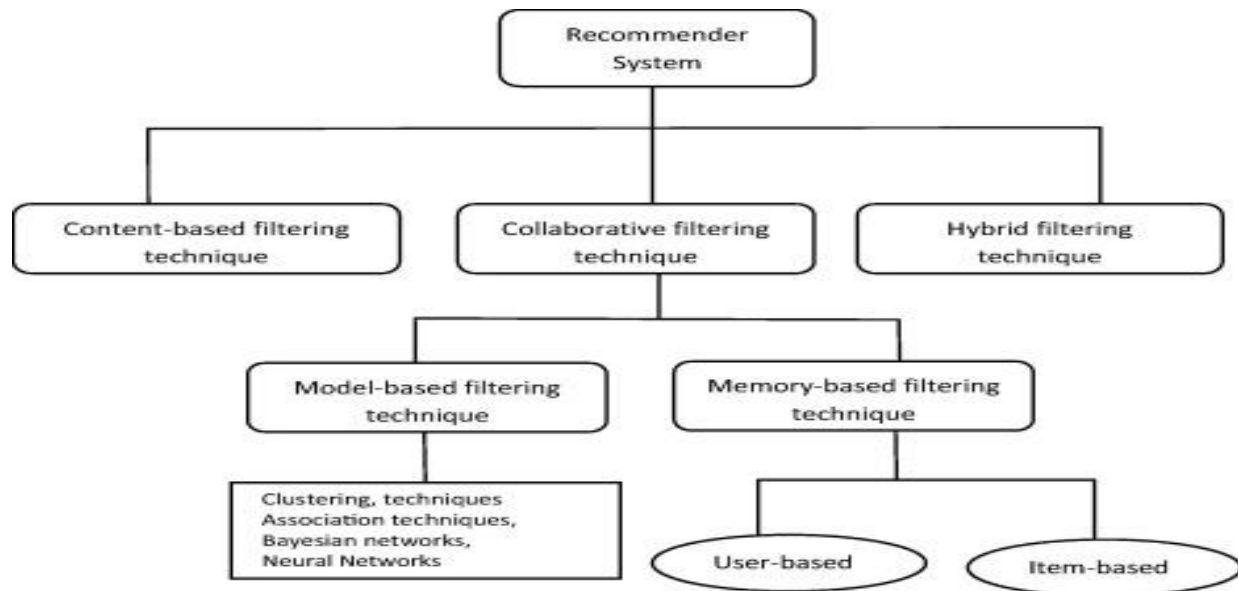


## MUSIC RECOMMENDATION SYSTEM

### Overview of project:

Recommender systems are beneficial to both the service providers and users. They Reduce the transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve decision making process and quality.

With the increase in number of people who are very much interested in listening to the music, it is highly advantage for the sites to introduce a recommendation technique which would recommend songs to the user which more likely to the user interest. Music recommendation system implemented by us would compute the similarities of the songs using several techniques called the Cosine similarity, Jaccard similarity and the Pearson coefficient and would recommend the most relevant songs to the user.

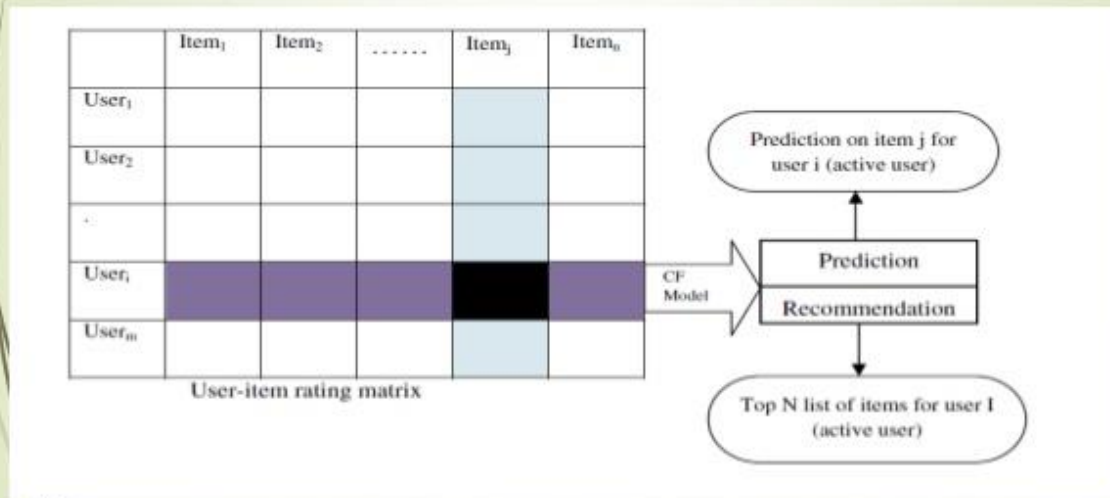


### Collaborative filtering technique:

We have used the most popular technique called the collaborative filtering which would help in recommending the items to the user ultimately. For every user recommender system would recommend items based on how similar users liked the item. considering an example for suppose Anusha and Chandrika have similar interests in music. Anusha recently listened to the song Despacito but Chandrika has never listened to this song since the system has learned that both have similar tastes in music it would recommend despacito to chandrika. In addition to this recommender systems can also perform collaborative filtering using the item similarity. Let's say that if one user is interested in despacito and liked it then the recommender system would suggest the other song to the user saying that many users also liked song2 along with despacito.

This is all done through finding similarity between user-user or item-item and then we calculate predicted ratings respectively and finally sort and get top 10 ratings. This can be done in 2 different ways User-based or Item-based Collaborative Filtering.

## Collaborative Filtering Process (Cont.)



In our project we have used the item based collaborative filtering technique to recommend songs to the user.

### **Item-Based Collaborative Filtering:**

This is a memory based filtering technique where we first find similarity of an item to every other item and in next step we use this similarity to predict ratings and the top K ratings are taken by sorting and recommended to user. We prefer item based collaborative filtering over user based collaborative filtering due its advantages.

## Item-Based Collaborative Filtering

	T1	T2	T3	T4	T5	T6	T7	T8
U1	4		3	4		2		5
U2		3	4	2	5		2	5
U3		4		3	?	2	2	4
U4	3		5		2	4		3
U5	2	5		2	2	4		2

### Pros:

Items are simpler, and they belong to same genre its very rare that that a song belong to 2 genres, where-as user's interest keeps on changing, so item similarity is more meaningful.

Item – Item similarity is once found, and when a new item comes in, we don't need to calculate whole model again. Sparsity problem it is very hard to find users who have rated same items

Collaborative approach doesn't require much user profile life content based approach.

Works for any kind of item. No feature selection needed.

### Cons:

Cold start problem to recommend we need to have enough number of users to find a match.

First rater problem it is very hard to recommend unrated songs.

### Dataset used:

<https://webscope.sandbox.yahoo.com/catalog.php?datatype=r&did=2>

This data set is of form **USERID SONGID RATINGS**

The data set we downloaded has 20 data which are 10 train and 10 test.

### Algorithms Implemented:

In this project we have implemented the cosine similarity algorithm and the Pearson correlation coefficient algorithm.

**Cosine similarity:** Cosine similarity is one of the most commonly used measure of similarity.

Given two vectors of attributes,  $i$  and  $j$ , the cosine similarity,  $\cos(\theta)$ , is represented using a dot product and magnitude as

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

Mathematical Representation:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Problem with cosine similarity: Treats the missing values as the 0 which is consider negative ratings

## PEARSON CORRELATION:

Pearson correlation coefficient is a statistical formula that measures the strength between the variables. This can be used as it resolves unknown rating value by replacing it with average value that is it normalize all ratings and replace unknown ratings with 0 and by this we can even take account of not liked songs like ratings below 3 on scale of (1to 5) and finally tough rater's problem too.

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

Mathematical representation:

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Where:

$N$	=	number of pairs of scores
$\sum xy$	=	sum of the products of paired scores
$\sum x$	=	sum of x scores
$\sum y$	=	sum of y scores
$\sum x^2$	=	sum of squared x scores
$\sum y^2$	=	sum of squared y scores

## Prediction Computation:

Prediction can be computed using average weighted sum

Mathematical Equation

Weighted Sum of Similar items for targeted item for an User	=	$\frac{\sum_{\text{all similar items}} S(i,N) * R(i,N)}{\sum_{\text{all similar items}} S(i,N)}$	where N is items rated by the User, for ex: {S(i,b)*4.5}
---	---	--	--

Task: Implementing a music recommender system using java map-reduce process.

## EXECUTION PROCESS:

Step1: Process data in to set of rated songs by user

For Example: There are songs S1, S2, S3, S4, S5, S6, S7, S8 and ratings R1, R2, R3, R4, R5,R6, R7, R8

user1 has listened to songs S1, S2, S3, S4 and rated R1, R2, R3, R4 respectively

All rated songs are grouped together into one row with userID as unique

userId1                      S1=R1, S2=R2, S3=R3, S4=R4

Step2: Similarity of each item to every other item is found sim (S1, S2), sim (S1, S3), sim (S1, S4)  
.....

Step3: for example, select user userid1 unrated songs will be S5, S6, S7, S8

Now we predict rating using weighted sum = [sim (S1, S5) \* R1 + sim (S2, S5) \* R2 + sim (S3, S5) \* R3 + sim (S4, S5) \* R5] / [sim (S1, S5) + sim (S2, S5) + sim (S3, S5) + sim (S4, S5)]

Similarly find for S6, S7, S8 are calculated

Step4: Recommendation is done here we need to sort the predictions and reverse them as we get result in ascending order after sorting. We can pick top K songs

In our project we collected sample data in above format and suggested all songs they would listen too.

Evaluation:

The evaluation of the recommendation systems done by calculating accuracy using Mean Absolute Error Method



## Outputs:

### Data Processing output

```
100008 ,23670-->1,12348-->1,21332-->5,108637-->1,107591-->5,94305-->1,78930-->5,56148-->4,52627-->5,28872-->2
100009 ,31324-->5,15108-->4,25742-->5,128104-->1,126237-->5,93680-->3,86486-->5,58399-->4,37098-->5,35006-->5
10001 ,26710-->5,12214-->5,130360-->5,110474-->5,89081-->5,72265-->5,61623-->5,60670-->3,41620-->5,41572-->5
100010 ,57180-->5,45606-->2,50683-->4,135051-->5,128261-->5,124153-->5,118541-->5,102568-->4,85585-->4,60191-->4
100011 ,30268-->4,19777-->3,21159-->3,131048-->5,88651-->5,67818-->5,54986-->2,52664-->5,43827-->5,37378-->4
100012 ,37400-->5,9003-->4,22955-->2,136388-->4,136283-->1,135510-->5,123036-->5,100966-->5,98625-->4,79954-->4
100013 ,56261-->3,14707-->3,44753-->4,131096-->3,115160-->4,95555-->3,77693-->2,64327-->3,62519-->5,61276-->2
100014 ,61318-->3,27226-->5,47162-->2,123220-->3,112452-->3,109887-->2,102830-->2,83821-->4,71398-->3,67612-->3
100015 ,43453-->5,3246-->5,42273-->5,125591-->5,119973-->5,110134-->5,92880-->5,67420-->5,64382-->5,46977-->5
100016 ,21054-->4,13890-->5,15527-->1,123506-->3,103402-->3,75987-->5,59186-->1,41347-->2,40660-->5,23314-->4
100017 ,28092-->5,13325-->1,22763-->5,136258-->2,131008-->2,110003-->2,87361-->2,84589-->1,33270-->5,28867-->2
100018 ,66931-->1,33664-->5,55107-->3,135176-->2,118514-->5,114087-->1,112214-->2,106292-->5,100605-->2,88606-->2
100019 ,41673-->2,19192-->5,28514-->5,133967-->1,127695-->1,123400-->1,122449-->3,118683-->3,61030-->2,60983-->3
10002 ,16276-->5,130435-->5,127437-->5,103603-->5,90915-->5,81143-->5,67437-->5,53676-->5,45842-->5,42090-->5
100020 ,37312-->2,2931-->3,37280-->5,124059-->3,123003-->5,116305-->2,82043-->3,79773-->4,59483-->4,45549-->2
100021 ,54109-->3,18184-->5,19116-->3,128445-->2,119914-->1,106785-->5,95524-->1,80562-->1,78747-->1,71550-->2
100022 ,12380-->5,3879-->4,8063-->3,134265-->1,125111-->3,102673-->3,88116-->4,64045-->5,46588-->1,45026-->1
100023 ,3724-->2,1048-->3,2217-->4,93459-->1,77048-->2,76952-->5,66818-->1,53917-->5,32469-->5,11953-->3
100024 ,62695-->5,799-->5,12264-->5,131761-->4,125197-->2,122710-->5,108430-->2,108345-->1,76432-->4,63963-->5
100025 ,18177-->5,3152-->5,14954-->1,117003-->1,107626-->4,68226-->1,57405-->2,55274-->1,21251-->3,19533-->2
100026 ,21849-->1,2713-->4,16794-->2,97414-->3,84409-->1,83368-->4,80551-->2,75215-->5,72762-->4,40037-->4
100027 ,11779-->3,3889-->5,4957-->5,112023-->5,93524-->4,89635-->5,82156-->1,73760-->4,66690-->5,37439-->2
100028 ,24464-->3,6719-->4,9450-->4,122192-->3,93703-->3,89984-->2,80062-->4,78887-->4,78452-->4,74958-->4
100029 ,49380-->3,12380-->1,14698-->4,136449-->2,128719-->2,109548-->5,73020-->5,67912-->2,60952-->5,53200-->3
10003 ,271-->5,130026-->5,113178-->5,105322-->5,63063-->5,54193-->3,48151-->5,47757-->5,18198-->3,6047-->5
100030 ,37206-->5,24407-->5,35406-->5,125042-->5,122684-->5,119009-->1,118832-->5,115038-->5,62686-->5,51521-->2
100031 ,14078-->1,6017-->1,9786-->3,116754-->3,99787-->1,93678-->5,74817-->3,47800-->1,42045-->5,34796-->3
100032 ,44466-->3,19172-->4,31306-->2,124746-->3,119778-->2,114054-->3,104960-->3,96261-->3,79323-->4,46790-->4
100033 ,16909-->5,5342-->5,11386-->1,130232-->5,126393-->5,119860-->5,111094-->5,79804-->1,78174-->5,42173-->5
100034 ,47991-->1,1048-->3,9186-->4,134620-->5,134265-->3,131570-->1,123590-->5,116763-->3,116522-->4,64942-->4
100035 ,36715-->1,2217-->4,10887-->2,128816-->1,110229-->4,107483-->2,78918-->1,53405-->3,52627-->2,49855-->3
100036 ,112495-->1,39223-->4,94604-->1,90562-->3,116608-->1,94247-->4,136383-->3,107544-->1,33584-->1,12861-->1
100037 ,112120-->5,2336-->5,119013-->5,106866-->2,16794-->3,107476-->1,128445-->5,10678-->2,105546-->5,82784-->5
100038 ,77376-->5,99802-->1,112306-->5,3418-->5,17910-->3,30507-->1,78755-->1,63031-->1,43799-->1,53252-->1
100039 ,22042-->5,7788-->3,114507-->5,71081-->4,53600-->3,84686-->5,22207-->5,108811-->2,131154-->2,54315-->3
10004 ,12994-->3,4637-->4,132122-->3,117363-->3,97306-->3,74131-->1,71016-->5,62636-->4,61298-->4,28941-->4
100040 ,84055-->5,29587-->3,88783-->5,13695-->2,40980-->4,61717-->4,67191-->2,36865-->3,46182-->5,126918-->5
100041 ,132633-->4,128647-->3,116756-->5,21391-->1,21542-->5,55712-->4,19333-->1,36582-->5,114330-->5,92195-->5
100042 ,122112-->5,74816-->1,111505-->1,28720-->5,31984-->4,63246-->1,22900-->1,23216-->1,69436-->5,99556-->5
100043 ,2783-->3,105432-->5,127413-->2,82029-->1,134387-->1,7614-->5,54255-->2,93218-->3,50215-->2,95575-->1
100044 ,111504-->2,119511-->2,129319-->3,74793-->3,22331-->3,58694-->4,125010-->4,79582-->4,44794-->2,39919-->5
100045 ,105421-->5,114848-->5,95868-->5,52049-->5,9280-->5,79847-->5,7889-->5,20527-->4,64312-->5,117089-->5
100046 ,56344-->1,24427-->3,94628-->2,59309-->1,74640-->1,75128-->1,79804-->1,105078-->1,130232-->1,132798-->1
100047 ,98841-->4,73892-->4,123791-->1,40712-->1,60211-->5,68151-->3,21110-->3,99556-->5,97004-->4,98093-->5
100048 ,22746-->5,45610-->4,129017-->5,88786-->5,60455-->3,12513-->5,34249-->3,125557-->5,54907-->5,78350-->2
100049 ,14623-->2,97564-->4,42351-->1,73453-->3,31102-->1,98988-->1,68002-->4,53084-->1,60816-->5,11386-->5
```

Similarity

100018  
104727=0. 9999999999999999, 114440=1. 0000000000000002, 119541=0. 9938837346736188, 119588=0. 9486832980505138, 120855=0. 8574929257125442, 123176=0. 9938837346736188, 123217=0. 9999999999999998, 123711=0. 9701425001453319, 124558=0. 9999999999999999!  
100028 133060=0. 827328794659617, 135711=1. 0, 135079=0. 7452413135250994  
100031 103109=0. 9909924304103233  
100032 113695=1. 0, 121425=0. 7657048647896112, 124533=0. 9970544855015815, 133717=0. 9941348467724341, 43267=0. 8320502943378437, 114848=0. 8320502943378437  
100033 103914=0. 9486832980505138  
100042 126757=0. 8944271909999159  
100056 117161=0. 9999999999999998, 121682=0. 9999999999999998  
100061 107002=1. 0000000000000002  
100069 105433=0. 5895494936611665, 134732=0. 9771398364036774, 103879=0. 8320502943378437, 117998=0. 9938837346736188, 108550=0. 9486832980505138  
10008 55061=0. 9037378388935388  
100089 126079=0. 8541985556144386  
100092 109557=0. 9682773237093576  
100095 120010=0. 9938837346736188, 129160=0. 5368754921931592  
100096 126282=0. 9999999999999999  
100098 119065=0. 9191450300180578, 113607=0. 9999999999999999, 105433=0. 9899494936611665, 106958=0. 9938837346736188, 122976=0. 9908681377244377, 107034=0. 9785497849867492  
100106 74249=0. 8320502943378437, 105784=0. 5462677805469224, 116183=0. 919145030018058, 129319=0. 9701425001453319  
100114 112576=0. 9938837346736188  
100116 114298=0. 9417419115948374, 105189=0. 8320502943378437  
100133 129323=0. 9922778767136677  
100148 125557=0. 9701425001453319  
10015 82156=0. 991117008018872  
100150 105421=0. 8320502943378437  
100152 14625=0. 9999999999999999, 126772=0. 9191450300180578, 128139=0. 9999999999999999, 129365=0. 9701425001453319, 133286=1. 0000000000000002, 70878=0. 8320502943378437, 75053=0. 9999999999999999, 114330=0. 8320502943378437  
100156 125626=0. 9999999999999999, 125557=0. 9725179925282851  
100162 120464=0. 9486832980505138  
100163 100584=0. 9079593845004517  
100178 110257=0. 9999999999999998  
100179 131814=0. 9701425001453319  
10018 57462=0. 994917930205513  
100192 108069=0. 9486832980505138, 114471=0. 8436614877321075, 118447=0. 9832820049844603  
100201 109267=0. 9938837346736188  
100206 102931=0. 9486832980505138, 132415=0. 9701425001453319  
10021 61364=0. 9761870601839526  
10022 45026=0. 9922778767136677  
100223  
135513=0. 9938837346736188, 105433=0. 9493496680149629, 105734=0. 9999999999999998, 108283=0. 5462677805469224, 113032=0. 9938837346736188, 110745=0. 8944271909999159, 127672=1. 0, 128445=1. 0, 128508=1. 0000000000000002, 134732=0. 8944271909999159, 131  
100224 106765=0. 9999999999999999  
100232 114899=0. 9486832980505138  
100233 106814=1. 0000000000000002, 126742=0. 8944271909999159, 127935=0. 9191450300180578, 128213=0. 9701425001453319  
10024 131963=0. 9260923597695477, 121933=0. 8320502943378437, 42906=0. 9701425001453319, 22593=0. 9682773237093576  
100247 109244=0. 9570244044334736, 115856=0. 7738413334506792, 60433=1. 0, 109151=0. 8320502943378437  
100253  
128718=0. 717137165006361, 101591=0. 9682773237093576, 102785=1. 0000000000000002, 104163=0. 888217643155949, 104457=0. 970142500145332, 105505=0. 8944271909999159, 110599=0. 9939123761133186, 111546=0. 9999999999999998, 112403=0. 9792166503011221, .'

Prediction

100008	12348	1	0
100008	21332	5	0
100008	23670	1	0
100008	28872	2	0
100008	52627	5	4
100008	56148	4	5
100008	78930	5	0
100008	94305	1	0
100008	107591	5	5
100008	108637	1	0
100009	15108	4	0
100009	25742	5	0
100009	31324	5	1
100009	35006	5	0
100009	37098	5	0
100009	58399	4	0
100009	86486	5	0
100009	93680	3	0
100009	126237	5	0
100009	128104	1	5
10001	12214	5	0
10001	26711	5	0
10001	41572	5	0
10001	41620	5	0
10001	60670	3	0
10001	61623	5	0
10001	72265	5	0
10001	89081	5	0
10001	110474	5	0
10001	130360	5	0
100010	45606	2	0
100010	50683	4	4
100010	57180	5	5
100010	60191	4	0
100010	85585	4	0
100010	102568	4	5
100010	118541	5	5
100010	124153	5	5
100010	128261	5	5
100010	135051	5	0
100011	19777	3	4
100011	21159	3	0
100011	30268	4	4
100011	37378	4	0
100011	43827	5	4
100011	52664	5	5



## Recommendation

```
100008 [ 92224, 66069, 83661, 73622, 128213, 64045, 135572, 51615, 71649, 71582]
100009 [ 82576, 37966, 41759, 105189, 83543, 52627, 85983, 121933, 116053, 116089]
10001 [ 60465, 70721, 135510, 132386, 68572, 111951, 67351, 52151, 22763, 22746]
100010 [ 80296, 106774, 61364, 90000, 122525, 134620, 6337, 126188, 124832, 118034]
100011 [ 66175, 101121, 52246, 108728, 126079, 61142, 12960, 100949, 136383, 91874]
100012 [ 86946, 106434, 14039, 47338, 126742, 53043, 84409, 76942, 127213, 109046]
100013 [ 132189, 119511, 120479, 117577, 121462, 129225, 48745, 124059, 130360, 136283]
100014 [ 101274, 79757, 129741, 50084, 128312, 108422, 83197, 78989, 122021, 2263]
100015 [ 63611, 111191, 91266, 126757, 79828, 59016, 123452, 132680, 83977, 88023]
100016 [ 111098, 50101, 133300, 80369, 93231, 109990, 76455, 53043, 56885, 118559]
100017 [ 118596, 83640, 118559, 132315, 130253, 56660, 122648, 118330, 70283, 70318]
100018 [ 80787, 113740, 67434, 87284, 92475, 40712, 132680, 63611, 134870, 34131]
100019 [ 124307, 17029, 22593, 133479, 133847, 125197, 120733, 126221, 131681, 128213]
10002 [ 132680, 83709, 56447, 18986, 134343, 96195, 36253, 8761, 88947, 52033]
100020 [ 50086, 41347, 129160, 70721, 53560, 44988, 91266, 107388, 89983, 37418]
100021 [ 123827, 134163, 132780, 113147, 128080, 113986, 128603, 126079, 136307, 11815]
100022 [ 78174, 95463, 18118, 114378, 79436, 22933, 127269, 122449, 122490, 101363]
100023 [ 78182, 120530, 99095, 78174, 89876, 66155, 90000, 116526, 108811, 105081]
100024 [ 53635, 82636, 52049, 75437, 37549, 102166, 93906]
100025 [ 118562, 52611, 129203, 6049, 111923, 28482, 110293, 119577, 77583, 96439]
100026 [ 109250, 106135, 80571, 88001, 101274, 112829, 78288, 101121, 87714, 132580]
100027 [ 118594, 120490, 132122, 118541, 94569, 128464, 5324, 12932, 26703, 97148]
100028 [ 131104, 34134, 27255, 96624, 25321, 30617, 129124, 129125, 28200, 132220]
100029 [ 83669, 118693, 126060, 76102, 134724, 124695, 85174, 117769, 115856, 89140]
10003 [ 112881, 132175, 89964, 32619, 136699, 25182, 119389, 32744, 94430, 17910]
100030 [ 124307, 132189, 128080, 72357, 120697, 67818, 54959, 65793, 111458, 130892]
100031 [ 75038, 126221, 112420, 95927, 104042, 80450, 125173, 51633, 135591, 131814]
100032 [ 78182, 132220, 112881, 97576, 105117, 89982, 132160, 68252, 33057, 66059]
100033 [ 91266, 26180, 23635, 54702, 118492, 100981, 96187, 126961, 135070, 81035]
100034 [ 131223, 34797, 135919, 103438, 101268, 16401, 102380, 54271, 9367, 71762]
100035 [ 85630, 49585, 69319, 101274, 8736, 45026, 62094, 125199, 61298, 61296]
100036 [ 126221, 132175, 126300, 70476, 134318, 134265, 95524, 134245, 112716, 52611]
100037 [ 80156, 85235, 79207, 125173, 135358, 131681, 112375, 56412, 85554, 130232]
```

## Status:

**DEFINITELY ACCOMPLISHED:** Our project will evaluate all the metrics used to find the similar items. Based on these ratings the songs are sorted and relevant songs are recommended using Item Based Collaborative filtering technique in MapReduce.

**LIKELY TO BE ACCOMPLISHED:** We will also consider various parameters while recommending songs such as artist, album, music quality etc.,

**IDEALLY TO BE ACCOMPLISHED:** We will try implement our code in python and also use ALS technique.

## Team Members:

Anusha Reddy Amula 800989169 [aamula@uncc.edu](mailto:aamula@uncc.edu)

Keerthi Sagar Bukkapuram 800954091 [kbukkapu@uncc.edu](mailto:kbukkapu@uncc.edu)

Guna Chandrika Paturi 800962470 [gpaturi@uncc.edu](mailto:gpaturi@uncc.edu)

## Task Division:

Learning phase: Data collection and gathering information: Anusha, Guna, Keerthi Sagar



Data Processing: Keerthi Sagar

Cosine Similarity: Keerthi Sagar

Pearson Similarity: Anusha

Prediction: Guna

Recommendation: sorting: Anusha, Guna

Webpages and proposal: Anusha, Guna, Keerthi Sagar

**Webpages:**

Report: [https://s3.amazonaws.com/bukkapu1/cloud\\_report.pdf](https://s3.amazonaws.com/bukkapu1/cloud_report.pdf)

Proposal: [https://s3.amazonaws.com/bukkapu1/Project\\_Proposal.pdf](https://s3.amazonaws.com/bukkapu1/Project_Proposal.pdf)

**References:**

[https://en.wikipedia.org/wiki/Recommender\\_system](https://en.wikipedia.org/wiki/Recommender_system)

<http://www.sciencedirect.com/science/article/pii/S1110866515000341>

[https://cseweb.ucsd.edu/~jmcauley/cse255/reports/wi15/Guanwen%20Yao\\_Lifeng\\_Cai.pdf](https://cseweb.ucsd.edu/~jmcauley/cse255/reports/wi15/Guanwen%20Yao_Lifeng_Cai.pdf)

<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7910613>

<https://www.youtube.com/watch?v=h9gpufJFF-0>

