Data Mining: Assignment 1

Group 12

Harshit Pant - CS21BTECH11021

Mahin Bansal - CS21BTECH11034

Satpute Aniket Tukaram - CS21BTECH11056

Vishal Vijay Devadiga - CS21BTECH11061

Question 1

Following are the top-3 answerers on the website and the top-3 tags with the most number of questions.

Rank	**User ID**	**Display Name**	**Answer Count**
1	9113	Doc Brown	2839
2	177980	Ewan	2326
3	1204	Robert Harvey	2043

Table 1: Top-3 answerers on the website

Rank	**TagId**	**TagName**	**Answer Count**
1	609	design	5162
2	249	c#	4931
3	79	java	4928

Table 2: Top-3 tags with the most number of questions

Question 2

We created a utililty matrix where the rows represents the users and the columns represents the tags (User x Tag)
The dimensions of the expert matrix is $\boxed{(1192,952)}$.

Question 3

- The summation value of the utility matrix is $\boxed{41221}$.
- The highest row sum of the utility matrix is $\boxed{1161}$.
- The highest column sum of the utility matrix is $\boxed{1394}$.

- For the test set we have considered the test set to be of width $(\lceil 0.15 * users \rceil, \lceil 0.15 * items \rceil)$
- The summation value of the train matrix is $\boxed{40565}$.
- The dimension of the test matrix is $\boxed{(168,\ 143)}$.
- The summation of the test matrix is $\boxed{656}$.

Question 4

Method	**Rating Prediction Function**	**Metric**	**N=2**	**N=3**	**N=5**
item-item	Simple average	RMSE	0.81	0.774	0.768
	Weighted average	RMSE	0.801	0.763	0.780
user-user	Simple average	RMSE	0.901	0.860	0.804
	Weighted average	RMSE	0.900	0.858	0.803

Table 3: RMSE values for different methods and rating prediction functions

Question 5

Method	**Metric**	**K=2**	**K=5**	**K=10**
Without Regularization	RMSE	0.696	0.742	0.864
$\lambda_p = 0.001,\ \lambda_q = 0.003$	RMSE	0.691	0.726	0.866
$\lambda_p = 0.05,\ \lambda_q = 0.05$	RMSE	0.692	0.742	0.848
$\lambda_p = 0.5,\ \lambda_q = 0.75$	RMSE	0.832	0.829	0.832

Table 4: RMSE values for different methods and regularization parameters

Question 6

Algorithm	**Algorithm**	**RMSE (N=2)**	**RMSE (N=3)**	**RMSE (N=5)**
item-item	Your method	0.81	0.774	0.768
	Surprise	0.9316	0.9204	0.8947
user-user	Your method	0.900	0.858	0.803
	Surprise	0.9492	0.9482	0.9399

Table 5: RMSE values for different algorithms

Method	**RMSE (K=2)**	**RMSE (K=5)**	**RMSE (K=10)**
Your method	0.691	0.726	0.832
Surprise	0.8402	0.8397	0.8393

Table 6: RMSE values for different algorithms