RESULTS

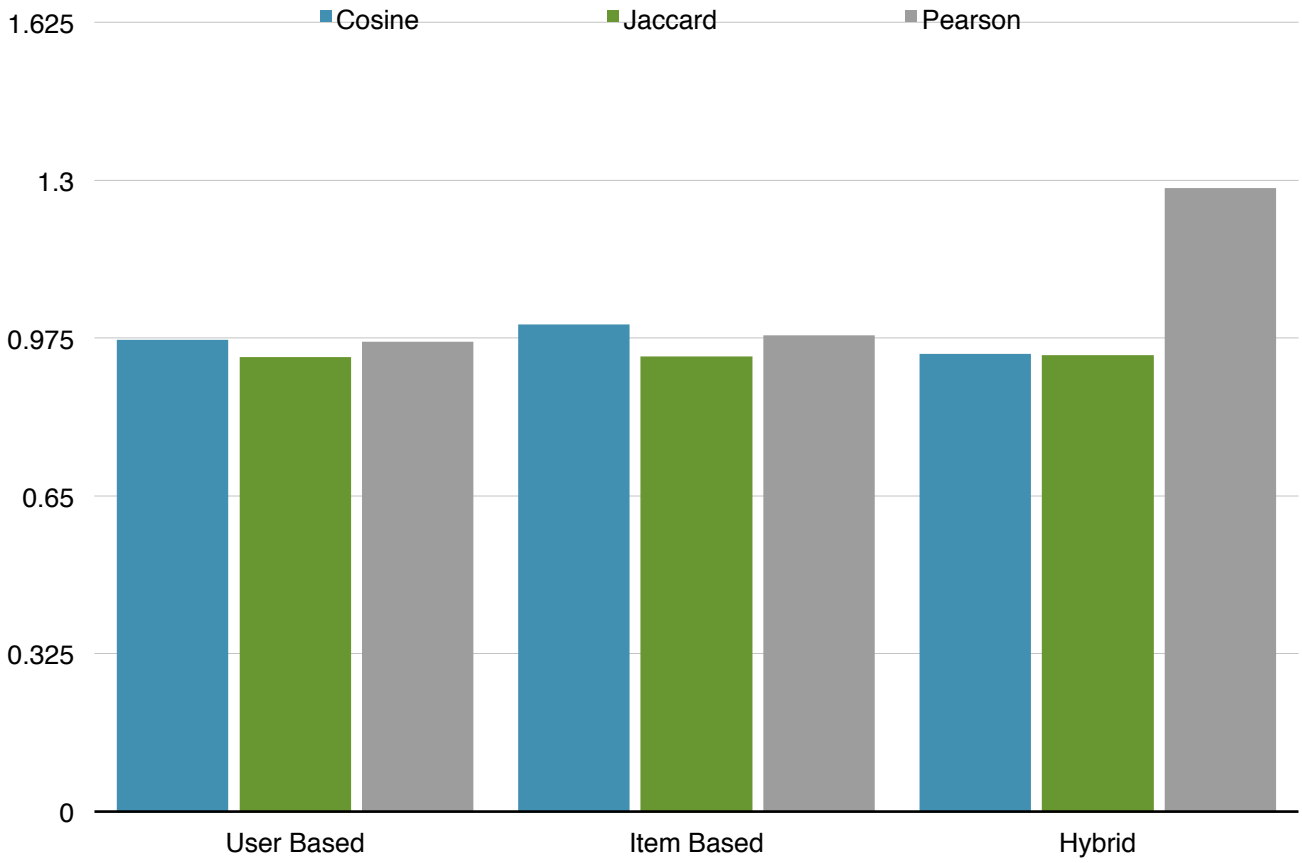

Fig : Bar Chart for Mean Squared Error for all the three types of similarities and for each of the user, item and hybrid based methods.

Part a:
The bar chart shows that the Jaccard similarity will give us the least Root mean square error for all the three types : Item-Based, User-Based and Hybrid. Also, we also see from the bar chart above that cosine similarity performs the least for User based and Item based whereas Pearson performs the worst for Hybrid-based recommender.

Part b:

For Result1 and Result2, I have reported them using the Jaccard similarity as the RMSE obtained after cross validation for them was of 0.9358 and 0.9365 respectively for the user based and item based recommendation. (It was lowest among the three similarties) The results1 and result2 contain exactly 200,000 entries that are the ratings predicted using the item based and user based method in the to-be-rated.csv file. I used multi fold cross validation to see which similarity out of the three gave the minimum RMSE and found that the Jaccard similarity worked out best for both the cases.

For Result3, I have used a hybrid of user-based method and item-based method. First, I will calculate all the three similarities between the users using the users.csv file and all the three similarities between the items using the movies.csv. Now, I will apply cross validation to predict which similarity results in the lowest rmse. I found that the lowest rmse was obtained using jaccard similarity. Now, to predict a rating, I used Jaccard

similarity to predict the rating for user and item based method, and then I took mean of the two ratings to obtain the final rating. This method ensures that we take into account the best rating obtainable from the user-based method and item based method. Lowest RMSE obtained in case of Hybrid is 0.9398.