

The following image show the output from the implementation of KNN with k in [8, 16, 32, 48, 64], for finding the optimal value of k in k-Nearest Neighbors. The accuracy metrics for each of the values of k

1. k = 8

```
In [13]: p, a = multi_k(8) # sachit: 16, hritik: 32
data = pd.DataFrame(list(zip(p, a)), columns=['pred', 'act'])
data.to_csv("recomm_results.csv", sep=',')
180436 of 183208 completed
180438 of 183208 completed
180440 of 183208 completed
180442 of 183208 completed
180445 of 183208 completed
180461 of 183208 completed
181784 of 183208 completed
181981 of 183208 completed
182144 of 183208 completed
182148 of 183208 completed
182299 of 183208 completed
182623 of 183208 completed
182628 of 183208 completed
182659 of 183208 completed
183113 of 183208 completed
183118 of 183208 completed
183133 of 183208 completed
183134 of 183208 completed
183140 of 183208 completed
8 - RMSE: 1.0295610274958815 - MAE: 0.7730702671400106 - R squared: 0.003431411382046168
```

2. k = 16

```
In [14]: p, a = multi_k(16)
data = pd.DataFrame(list(zip(p, a)), columns=['pred', 'act'])
data.to_csv("recomm_results.csv", sep=',')
180436 of 183208 completed
180438 of 183208 completed
180440 of 183208 completed
180442 of 183208 completed
180445 of 183208 completed
180461 of 183208 completed
181784 of 183208 completed
181981 of 183208 completed
182144 of 183208 completed
182148 of 183208 completed
182299 of 183208 completed
182623 of 183208 completed
182628 of 183208 completed
182659 of 183208 completed
183113 of 183208 completed
183118 of 183208 completed
183133 of 183208 completed
183134 of 183208 completed
183140 of 183208 completed
16 - RMSE: 0.9637281131512011 - MAE: 0.7278820873933527 - R squared: 0.12680334939330307
```

3. k = 32

```
In [13]: p, a = multi_k(32) # sachit: 16, hritik: 32

data = pd.DataFrame(list(zip(p, a)), columns=['pred', 'act'])
data.to_csv("recomm_results.csv", sep=',')

182299 of 183208 completed
--- 0.0006489753723144531 seconds ---
182623 of 183208 completed
--- 0.0005919933319091797 seconds ---
182628 of 183208 completed
--- 0.0005919933319091797 seconds ---
182659 of 183208 completed
--- 0.0006649494171142578 seconds ---
183113 of 183208 completed
--- 0.0005540847778320312 seconds ---
183118 of 183208 completed
--- 0.0006220340728759766 seconds ---
183133 of 183208 completed
--- 0.0005717277526855469 seconds ---
183134 of 183208 completed
--- 0.0007700920104980469 seconds ---
183140 of 183208 completed
--- 0.0006158351898193359 seconds ---
32 - RMSE: 0.9317721480593276 - MAE: 0.7073213882638242 - R squared: 0.18375139074258917
```

4. k = 48

```
In [14]: p, a = multi_k(48)

data = pd.DataFrame(list(zip(p, a)), columns=['pred', 'act'])
data.to_csv("recomm_results.csv", sep=',')

180438 of 183208 completed
180440 of 183208 completed
180442 of 183208 completed
180445 of 183208 completed
180461 of 183208 completed
181784 of 183208 completed
181981 of 183208 completed
182144 of 183208 completed
182148 of 183208 completed
182299 of 183208 completed
182623 of 183208 completed
182628 of 183208 completed
182659 of 183208 completed
183113 of 183208 completed
183118 of 183208 completed
183133 of 183208 completed
183134 of 183208 completed
183140 of 183208 completed
48 - RMSE: 0.924310909386484 - MAE: 0.7045150489189186 - R squared: 0.196771401530346
```

5. k = 64

```
In [13]: p, a = multi_k(64)

data = pd.DataFrame(list(zip(p, a)), columns=['pred', 'act'])
data.to_csv("recomm_results.csv", sep=',')

180436 of 183208 completed
180438 of 183208 completed
180440 of 183208 completed
180442 of 183208 completed
180445 of 183208 completed
180461 of 183208 completed
181784 of 183208 completed
181981 of 183208 completed
182144 of 183208 completed
182148 of 183208 completed
182299 of 183208 completed
182623 of 183208 completed
182628 of 183208 completed
182659 of 183208 completed
183113 of 183208 completed
183118 of 183208 completed
183133 of 183208 completed
183134 of 183208 completed
183140 of 183208 completed
64 - RMSE: 0.9221621378162773 - MAE: 0.7052846920683487 - R squared: 0.20050163682552857
```