In [1]:

```
import numpy as np
import pyarrow.feather as feather
import pandas as pd
from surprise.model_selection import train_test_split, GridSearchCV, RandomizedSearc
from joblib import parallel_backend
from surprise import KNNBasic
from sklearn.metrics import mean_squared_error
from surprise import Dataset, Reader, BaselineOnly, accuracy
```

In [2]:

```
df_train = feather.read_feather('netflix-5k.train.feather')
df_val = feather.read_feather('netflix-5k.validation.feather')
df_titles = feather.read_feather('netflix-5k.movie_titles.feather')
```

In [3]:

```
reader = Reader(rating_scale=(0, 5))
data = Dataset.load_from_df(df_train[['userID','movieID',
                                      'rating']],reader)
datav = Dataset.load_from_df(df_val[['userID','movieID',
                                     'rating']],reader)
trainset = data.build_full_trainset()
NA,valset = train_test_split(datav, test_size=1.0)
```

In [4]:

```
param_grid =  {'sim_options' : {'name': ['msd','pearson', 'pearson_baseline', 'cosir
                                'user_based': [False, True],
                                'shrinkage': [50, 75, 100, 125, 150],
                                'min_support': [2, 4, 6, 8, 10]
                               },
               'min_k': [1, 3, 5, 7, 9],
               'k': [5, 10, 30, 40, 50]
              }

with parallel_backend('multiprocessing', n_jobs=6):
    sim_options = RandomizedSearchCV(KNNBasic, param_grid, measures=['rmse'], cv=5,
                            n_jobs=6, joblib_verbose=10)
    sim_options.fit(data)

print("Best Score from Grid Search is ", sim_options.best_score['rmse'])
print("Best parameters for sim options for KNN Basic are", sim_options.best_params['
```

```
[Parallel(n_jobs=6)]: Using backend MultiprocessingBackend with 6 conc
urrent workers.
[Parallel(n_jobs=6)]: Done   1 tasks      | elapsed: 10.2min
[Parallel(n_jobs=6)]: Done   6 tasks      | elapsed: 22.6min
[Parallel(n_jobs=6)]: Done  13 tasks      | elapsed: 41.3min
[Parallel(n_jobs=6)]: Done  20 tasks      | elapsed: 48.3min
[Parallel(n_jobs=6)]: Done  29 tasks      | elapsed: 68.9min
[Parallel(n_jobs=6)]: Done  34 out of  40 | elapsed: 79.0min remainin
g: 13.9min

Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Estimating biases using als...
Computing the pearson_baseline similarity matrix...
Done computing similarity matrix.
Estimating biases using als...
Computing the pearson_baseline similarity matrix...
Done computing similarity matrix.
Estimating biases using als...
Computing the pearson_baseline similarity matrix...
Done computing similarity matrix.
Estimating biases using als...
Computing the pearson_baseline similarity matrix...
Done computing similarity matrix.
Estimating biases using als...
Computing the pearson_baseline similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
```

```
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the cosine similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Computing the pearson similarity matrix...
Done computing similarity matrix.
Best Score from Grid Search is  0.9008883835678553
Best parameters for sim options for KNN Basic are {'sim_options': {'na
me': 'pearson_baseline', 'user_based': False, 'shrinkage': 150, 'min_s
upport': 2}, 'min_k': 1, 'k': 40}

[Parallel(n_jobs=6)]: Done  40 out of  40 | elapsed: 84.6min finished
```

In [ ]: