**Description**

**Azamat Ordabekov**

1. Spark version: 2.3.1. Scala version: 2.11
2. In order to run my program, you need to put .jar file to spark’s bin folder and type next command:

**For model based:**

**spark-submit --class ModelBasedCF Azamat\_Ordabekov\_hw2.jar <train\_review\_path> <test\_review\_path>**

**For item based:**

**spark-submit --class ItemBasedCF Azamat\_Ordabekov\_hw2.jar <train\_review\_path> <test\_review\_path>**

Note: the path should be absolute.

1. The result of running my program:

**Model Based:**

>=0 and <1: 28500

>=1 and <2: 12187

>=2 and <3: 2238

>=3 and <4: 328

>=4: 34

RMSE: 1.0669456808189848

Time: 155.616233899 sec

**Item Based:**

>=0 and <1: 28954

>=1 and <2: 12849

>=2 and <3: 2716

>=3 and <4: 628

>=4: 89

RMSE: 1.1004433791369959

Time: 160.3771882 sec

1. **Improvements**
   1. Model-Based: I implemented Spark MLib. At the prediction, if the output is more than 5.00, set to equal 5.00. If the output is less than 1.00, set to equal 1.00.
   2. Item-Based: set neighborhood to 90. This neighborhood gave RMSE less than threshold. The problem of cold start, I solved by getting average of user stars. If user doesn’t exist in the train data, than set average business stars. If both user and business are new, then set prediction to 3.00. Additionally, if the output is more than 5.00, set to equal 5.00. If the output is less than 1.00, set to equal 1.00.