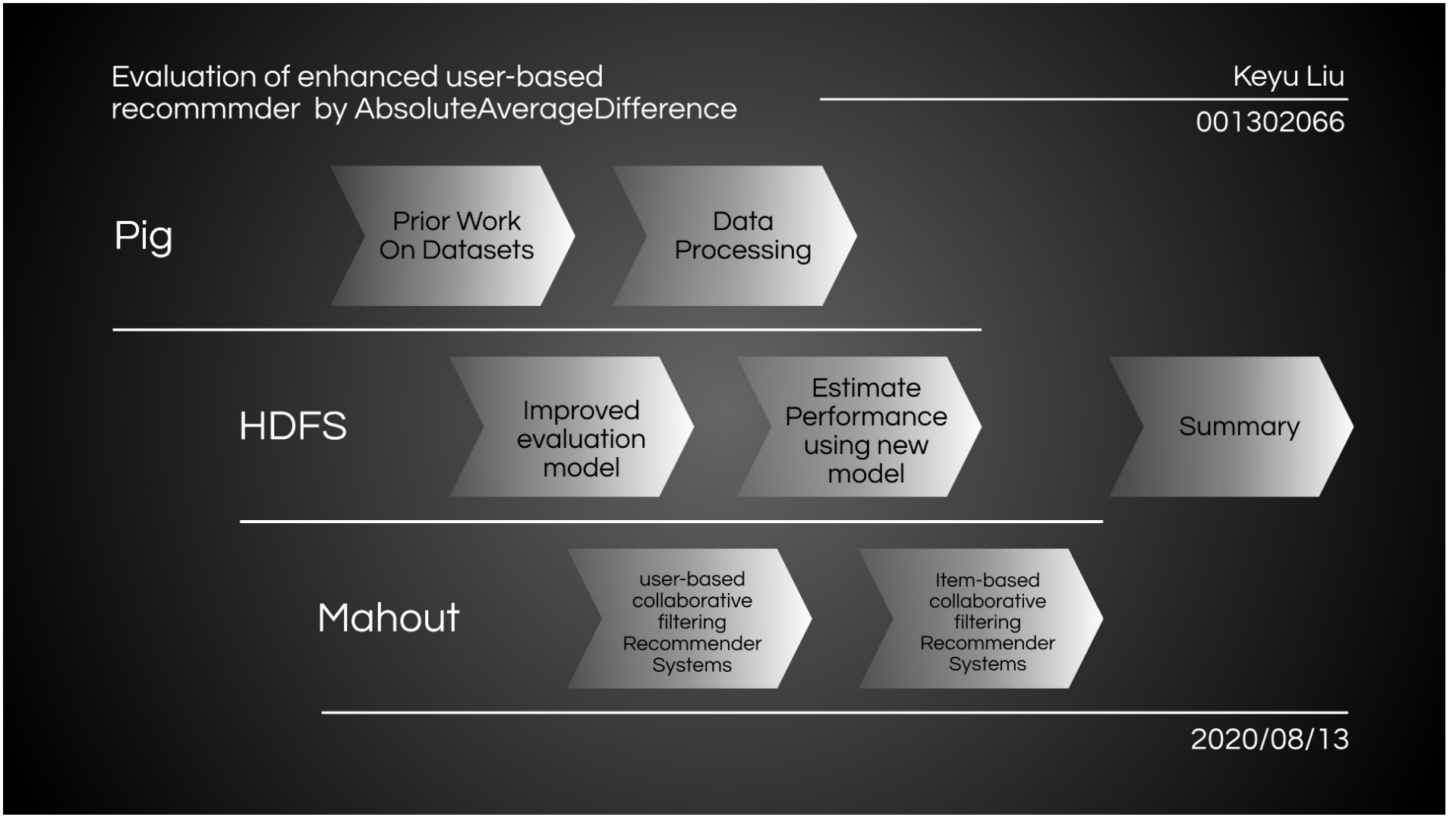
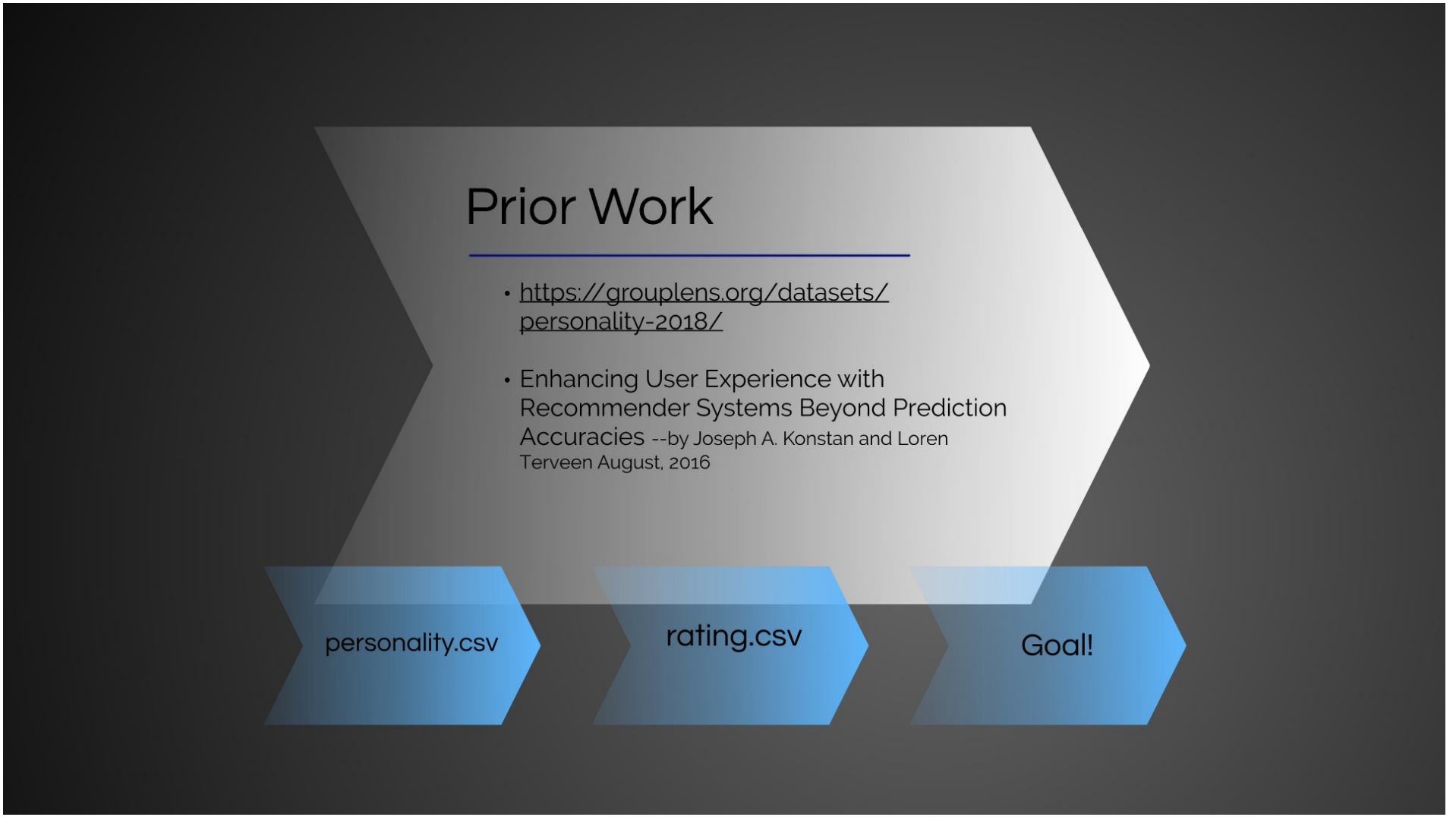
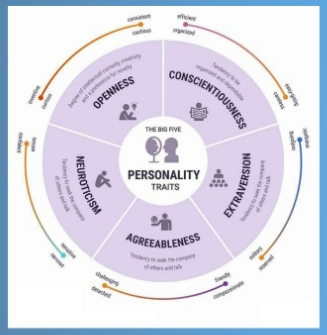
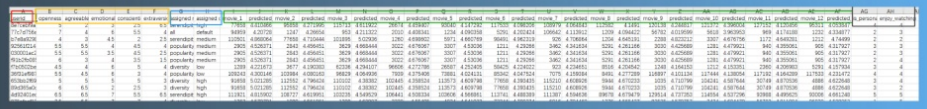

Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

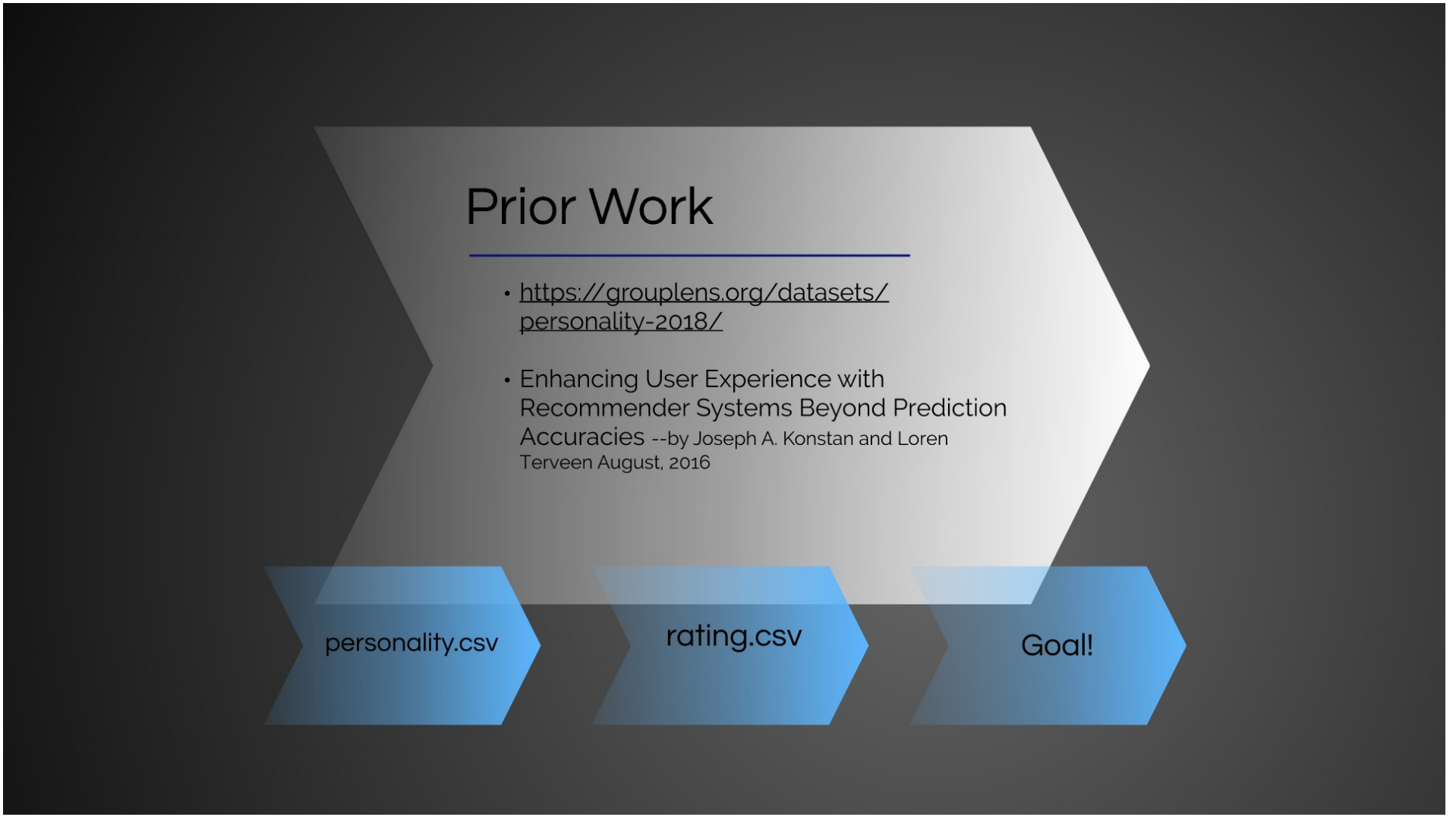

Prior Work
https://grouplens.org/datasets/personality-2018/
Enhancing User Experience with Recommender Systems Beyond Prediction Accuracies --by Joseph A. Konstan and Loren Terveen August, 2016
personality.csv
rating.csv
Goal!

persenality-data.csv

1. User ID
2. User_personality: OCEAN five character model
3. Predicted Result: 12 movies and predications
4. Evaluation for Prediction: is_personalized and enjoy_watching



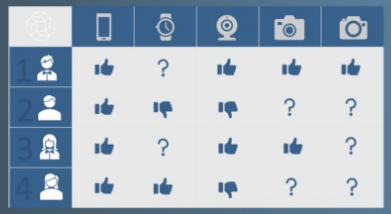


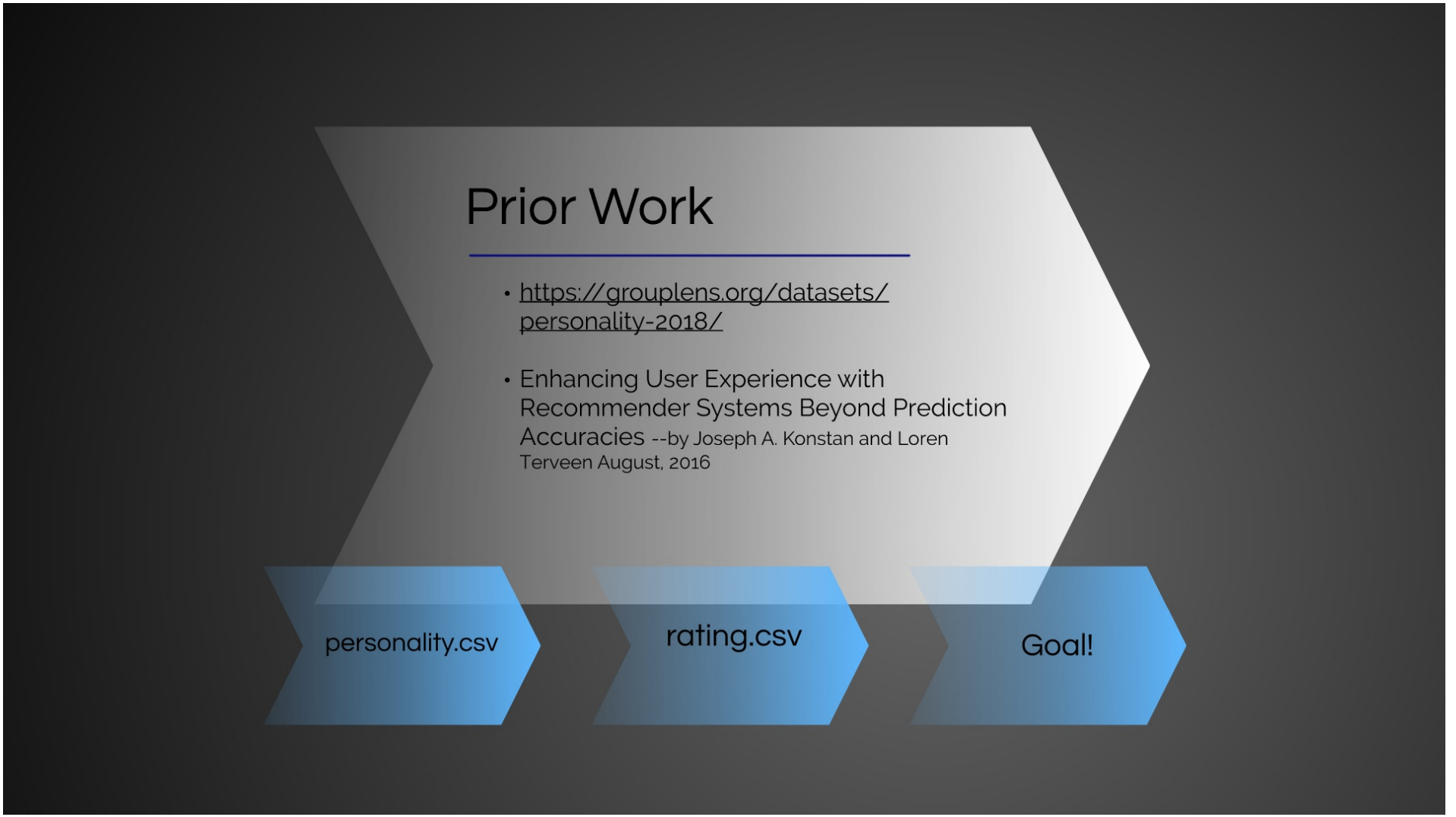

Prior Work
https://grouplens.org/datasets/personality-2018/
Enhancing User Experience with Recommender Systems Beyond Prediction Accuracies --by Joseph A. Konstan and Loren Terveen August, 2016
personality.csv
rating.csv
Goal!

rating.csv

1. User ID
2. Movie ID
3. Rating
4. Time Stamp

	A	B	C	D	E
1	useri	movie_id	rating	tstamp	
2	8e7cebf9a2	1	5	2001-09-10 17:19:56	
3	8e7cebf9a2	2	4	2001-09-28 11:34:55	
4	8e7cebf9a2	3	4	2001-09-28 11:42:50	
5	8e7cebf9a2	5	5	2001-09-28 11:27:30	
6	8e7cebf9a2	6	4	2002-01-07 18:12:02	
7	8e7cebf9a2	7	4	2001-09-21 14:09:24	
8	8e7cebf9a2	10	4	2001-09-29 18:13:15	
9	8e7cebf9a2	11	4	2001-09-21 11:54:21	
10	8e7cebf9a2	12	4	2001-09-29 18:58:15	
11	8e7cebf9a2	13	4	2001-09-14 15:29:46	

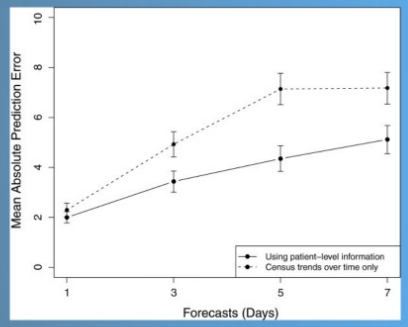


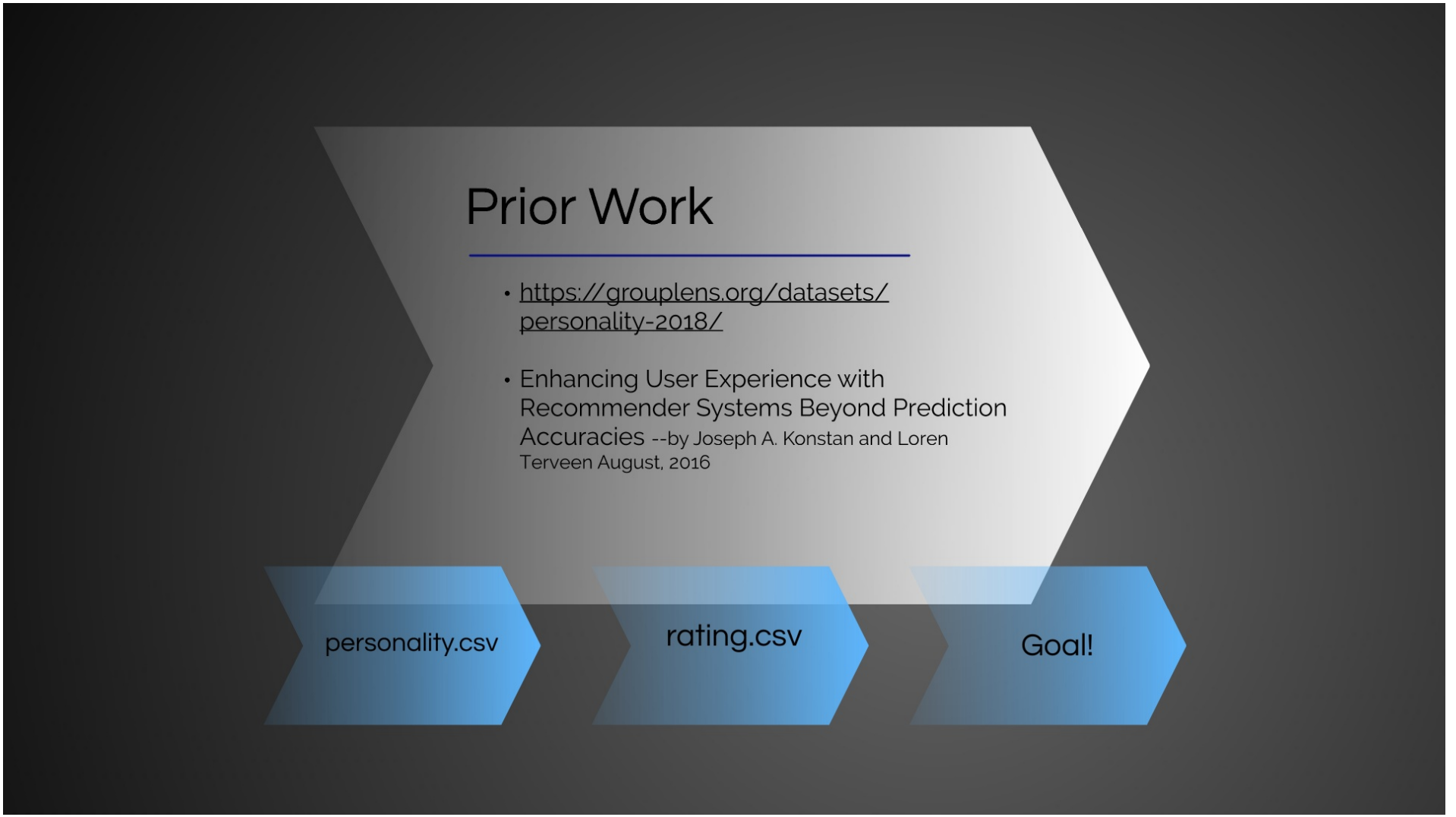

Prior Work
https://grouplens.org/datasets/personality-2018/
Enhancing User Experience with Recommender Systems Beyond Prediction Accuracies --by Joseph A. Konstan and Loren Terveen August, 2016
personality.csv
rating.csv
Goal!

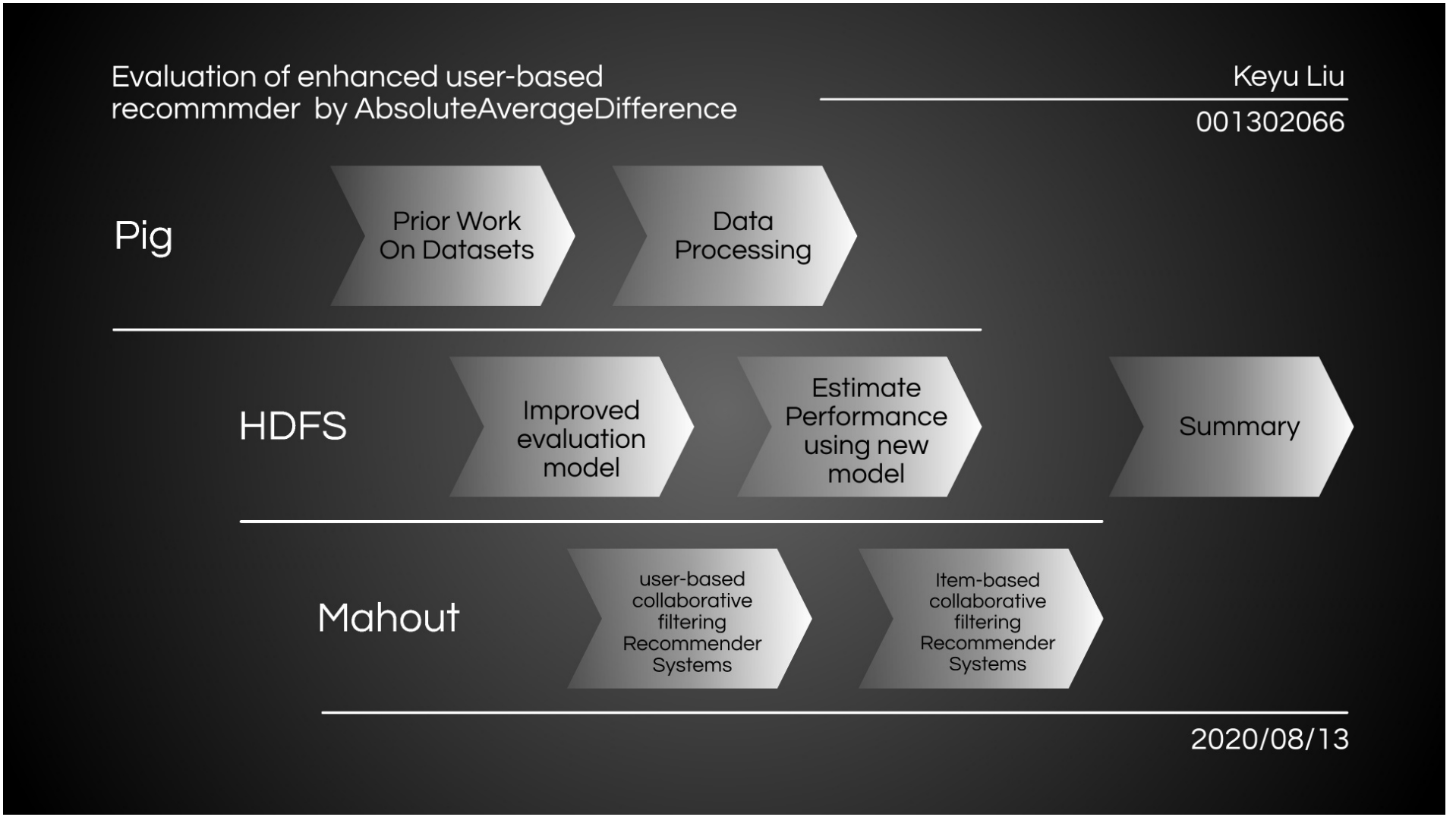
AbsoluteAverageDifference

- I will evaluate the performance of this enhanced user-based recommender by the absolute average difference between system recognition and user's feedback.

Why not use the predication data of 12 movies?




Prior Work
https://grouplens.org/datasets/personality-2018/
Enhancing User Experience with Recommender Systems Beyond Prediction Accuracies --by Joseph A. Konstan and Loren Terveen August, 2016
personality.csv
rating.csv
Goal!


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

Pig operation

- Obtain Original data from questionnaire - researchData.csv
- Average value of each users' (is_personalized) - is_personaized.csv
- Average value of each users' (enjoy_watching) - enjoy_watching.csv

Assume each user in the survey offers fair and reasonable evaluation.
we have the AbsoluteAverageDifference is 3.606783 - 3.524451 = **0.082332**

Join personalituy.csv and table.csv

MapReduce(userid, is_personalized)

MapReduce(userid, enjoy_watchingg

Join operation

reasearchData = JOIN personality BY userid, rating BY useri;

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.7.3   0.17.0  keyulu 2020-08-12 09:14:32     2020-08-12 09:14:49     HASH_JOIN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feat
job_local431138236_0016 3       1       n/a     n/a     n/a     n/a     n/a     n/a     n/a     n/a     personality,rating,reasearchData         HASH_JOIN       file:///home

Input(s):
Successfully read 1835 records from: "file:///home/keyulu/Downloads/pig-0.17.0/bin/personality.csv"
Successfully read 1028752 records from: "file:///home/keyulu/Downloads/pig-0.17.0/bin/ratings.csv"

Output(s):
Successfully stored 1070203 records in: "file:///home/keyulu/Downloads/pig-0.17.0/bin/output/researchData"

Counters:
Total records written : 1070203
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local431138236_0016


2020-08-12 09:14:49,139 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2020-08-12 09:14:49,139 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2020-08-12 09:14:49,139 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2020-08-12 09:14:49,141 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

```
(005fe8678214011d7f92e51f9a546d40, 5.5, 6.0, 2.0, 4.5, 1.0, popularity, low, 94466, 4.17266086321, 77658, 4.19966847544, 92259, 4.14083397435, 112556, 4
 4.12552693679, 3435, 4.22468194276, 44555, 4.23534266643, 27773, 4.09204931497, 1, 3 ,005fe8678214011d7f92e51f9a546d40, 1, 4, 2015-05-18 21:52:01 )
(005fe8678214011d7f92e51f9a546d40, 5.5, 6.0, 2.0, 4.5, 1.0, popularity, low, 94466, 4.17266086321, 77658, 4.19966847544, 92259, 4.14083397435, 112556, 4
 4.12552693679, 3435, 4.22468194276, 44555, 4.23534266643, 27773, 4.09204931497, 1, 3 ,005fe8678214011d7f92e51f9a546d40, 50, 4, 2015-05-19 17:58:26 )
```

```
| reasearchData     | personality::userid:bytearray       | personality::openness:bytearray       | personality::a
rray       | personality::assignedmetric:bytearray       | personality::assignedcondition:bytearray       | pers
array       | personality::movie_3:bytearray       | personality::predicted_rating_3:bytearray       | personality
     | personality::movie_6:bytearray       | personality::predicted_rating_6:bytearray       | personality::movie
personality::movie_9:bytearray       | personality::predicted_rating_9:bytearray       | personality::movie_10:byt
sonality::movie_12:bytearray       | personality::predicted_rating_12:bytearray       | personality::is_personaliz
y       | rating::tstamp:bytearray       |
```

Pig operation

- Obtain Original data from questionnaire - researchData.csv
- Average value of each users' (is_personalized) - is_personaized.csv
- Average value of each users' (enjoy_watching) - enjoy_watching.csv

Assume each user in the survey offers fair and reasonable evaluation.
we have the AbsoluteAverageDifference is 3.606783 - 3.524451 = **0.082332**

Join personalituy.csv and table.csv

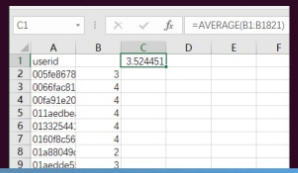
MapReduce(userid, is_personalized)

MapReduce(userid, enjoy_watchingg

enjoy_watching.csv

```
grpd = GROUP userinfo BY userid;
enjoy_watching  = FOREACH grpd GENERATE group, AVG(userinfo.enjoy_watching);
STORE enjoy_watching INTO 'output/enjoy_watching' USING PigStorage(',');
```

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.7.3   0.17.0  keyuliu 2020-08-12 12:47:17      2020-08-12 12:47:26     GROUP_BY

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinRedu
job_local362269726_0001 3       1       n/a     n/a     n/a     n/a     n/a     n/a     n/a     n/a     enjoy_w

Input(s):
Successfully read 1070204 records from: "file:///home/keyuliu/Downloads/pig-0.17.0/bin/userinfo.csv"

Output(s):
Successfully stored 1821 records in: "file:///home/keyuliu/Downloads/pig-0.17.0/bin/output/enjoy_watching"

Counters:
Total records written : 1821
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local362269726_0001
```



Pig operation

- Obtain Original data from questionnaire - researchData.csv
- Average value of each users' (is_personalized) - is_personaized.csv
- Average value of each users' (enjoy_watching) - enjoy_watching.csv

Assume each user in the survey offers fair and reasonable evaluation.
we have the AbsoluteAverageDifference is 3.606783 - 3.524451 = **0.082332**

Join personalituy.csv and table.csv

MapReduce(userid, is_personalized)

MapReduce(userid, enjoy_watchingg

rating.csv

```
grpd = GROUP userinfo BY userid;
rating =  FOREACH grpd GENERATE group, AVG(userinfo.rating);
STORE  rating INTO 'output/userinfo.rating' USING PigStorage(',');
```

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.7.3   0.17.0  keyuliu 2020-08-12 12:47:17      2020-08-12 12:47:26     GROUP_BY

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinRe
job_local362269726_0001 3       1       n/a     n/a     n/a     n/a     n/a     n/a     n/a     n/a     enjoy

Input(s):
Successfully read 1070204 records from: "file:///home/keyuliu/Downloads/pig-0.17.0/bin/userinfo.csv"

Output(s):
Successfully stored 1821 records in: "file:///home/keyuliu/Downloads/pig-0.17.0/bin/output/enjoy_watching"

Counters:
Total records written : 1821
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local362269726_0001
```

	A	B	C
1	userid		3.606783
2	005fe8678	4.111111	
3	0066fac81	3.971545	
4	00fa91e20	3.327285	
5	011aedbe	3.387879	

Pig operation

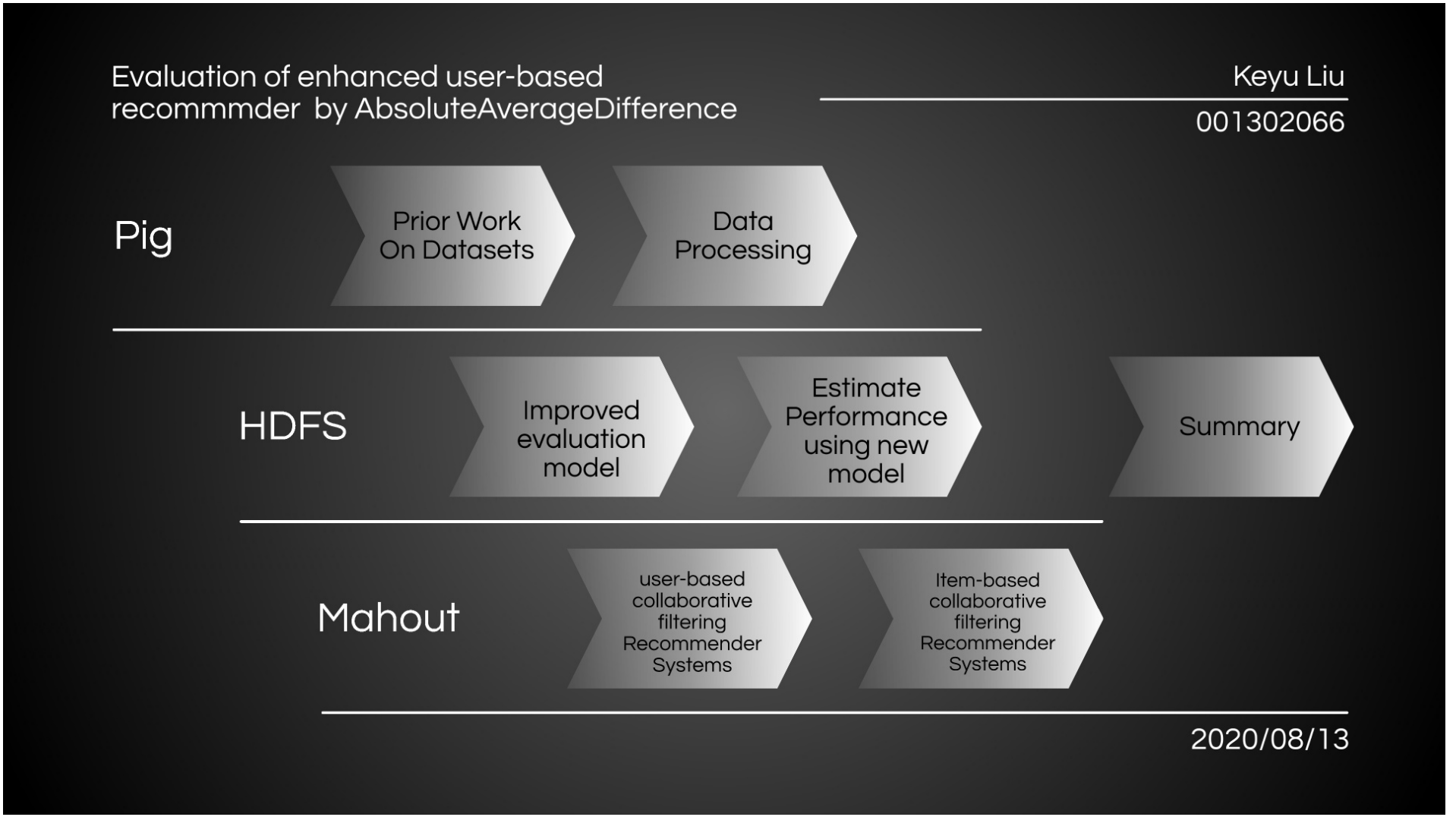
- Obtain Original data from questionnaire - researchData.csv
- Average value of each users' (is_personalized) - is_personaized.csv
- Average value of each users' (enjoy_watching) - enjoy_watching.csv

Assume each user in the survey offers fair and reasonable evaluation.
we have the AbsoluteAverageDifference is 3.606783 - 3.524451 = **0.082332**

Join personalituy.csv and table.csv

MapReduce(userid, is_personalized)

MapReduce(userid, enjoy_watchingg


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

Limitation

In fact, people with different personality will give different comment for same things even they have similarity taste.

1. Separate people by personality

2.Predicted Result Symbol

3. Realistic FeedBack Symbol

Using Persenality To Distinguish User

In MapReduce, the OCEAN five factor will be used to decide how partitioner works.

userid	openness	agreeablen	emotional_	conscientic	extraversic
8e7cebf9a2	5	2	3	2.5	6.5

Limitation

In fact, people with different personality will give different comment for same things even they have similarity taste.

1. Separate people by personality

2.Predicted Result Symbol

3. Realistic FeedBack Symbol

assignedcondition

As we have used enjoy_watching as predicted result from this enhanced user-based recommender system
Now the assignedcondition (high, midterm, low) will be assumed as the objective benchmark for movies.

To transfer these new benckmark to numbers in a resonable way. I will use quartering.

For example, there are totally 11 numbers from 0 to 10.
the value of
High = (11+1) * 0.75
miderm = (1+1) * 0.5
low + (11+1) * 0.25

Limitation

In fact, people with different personality will give different comment for same things even they have similarity taste.

1. Separate people by personality

2.Predicted Result Symbol

3. Realistic FeedBack Symbol

a*is_personalized +b*enjoy_watching
(a+b=2)

is_personalized +enjoy_watching will represent the feedback from user from they used this recommender system.

As is_personalized is from (1-5) and enjoy_watching is also from (1-5) .
And they need to sub the predicted result-assigned condition. So in the model I decide to multiple the assigned condition to make them balanced when calculating absolute average difference.
The representation of realistic feedback will be:

a*is_personalized +b*enjoy_watching (a+b=2)

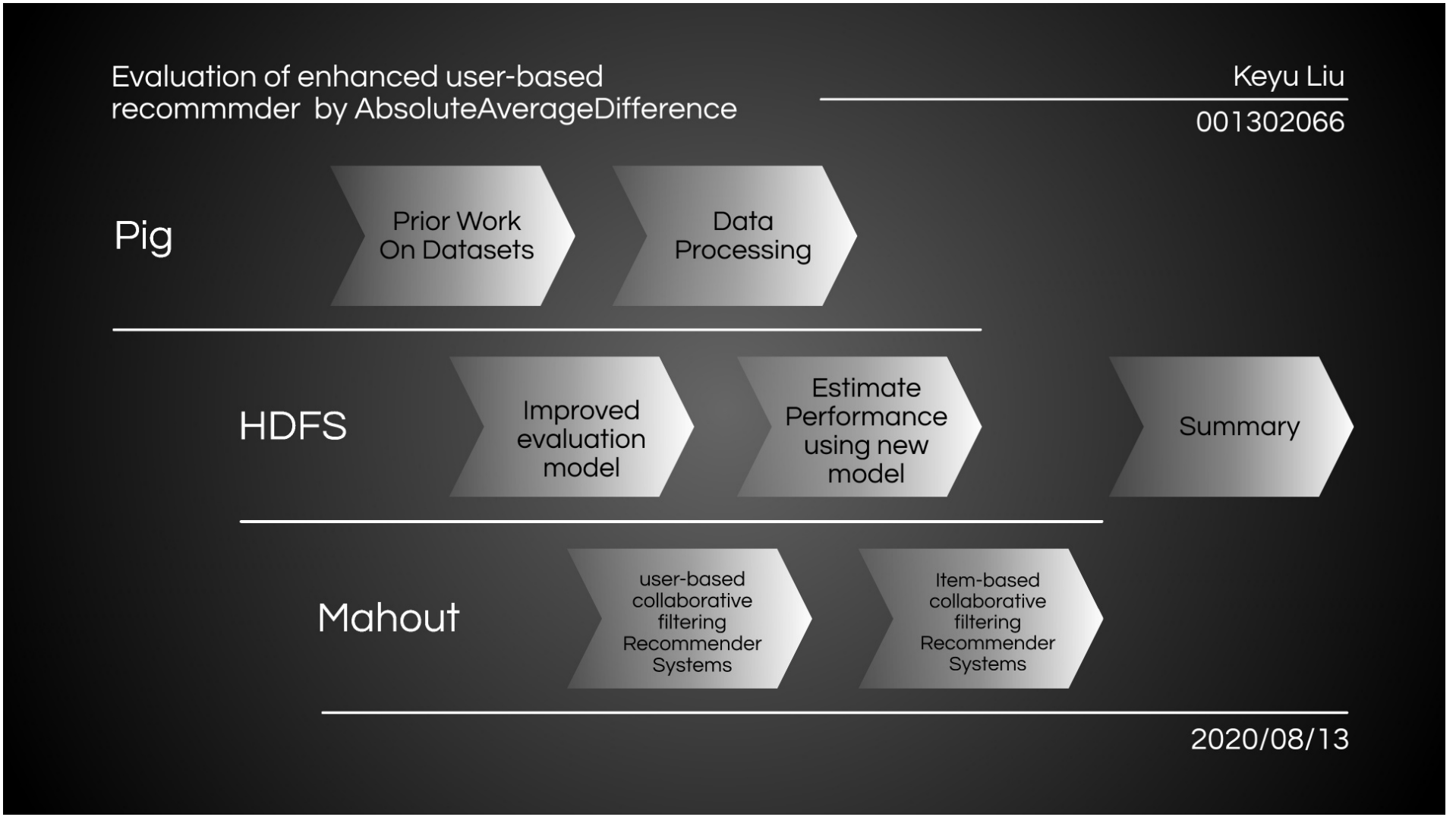
Limitation

In fact, people with different personality will give different comment for same things even they have similarity taste.

1. Separate people by personality

2.Predicted Result Symbol

3. Realistic FeedBack Symbol


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

Map Reduce

Discrete Function of AAD and personality

openness	agreeableness	emotional_stability	conscientiousness	extraversion		counter	absoluteDifference	averageAbsoluteDifference	1070203
20	30	40	60	15		1537	15370	10	
35	35	25	65	15		84	840	10	
45	35	50	40	15		1007	10070	10	
50	35	45	30	25		688	6880	10	
55	40	65	25	35		242	2420	10	
55	50	30	50	40		67	670	10	
55	50	65	55	55		1210	12100	10	
55	55	45	35	30		591	5910	10	
55	70	50	60	10		357	3570	10	
65	45	60	40	60		488	4880	10	

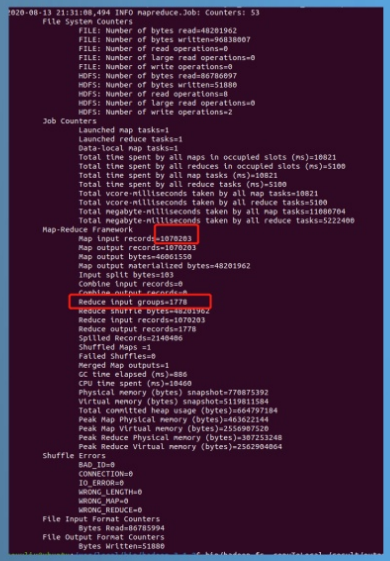
Converting to PDF using distributed linear algebra or knn/SVM. Then, building a model to analysis the relationship between personality and AAD.

ScreenShoot

bin/hadoop fs -mkdir /
EvaluationModel

bin/hadoop fs -copyFromLocal ~/
Downloads/userinfo.csv /input

bin/hadoop jar ~/Downloads/
SecondarySort-1.0-SNAPSHOT.jar
DriverClass /input/userinfo.csv /
result/output/Evaluation
bin/hadoop fs -copyToLocal /result/
output/Evaluation/part-r-00000 ~/
Downloads/Project

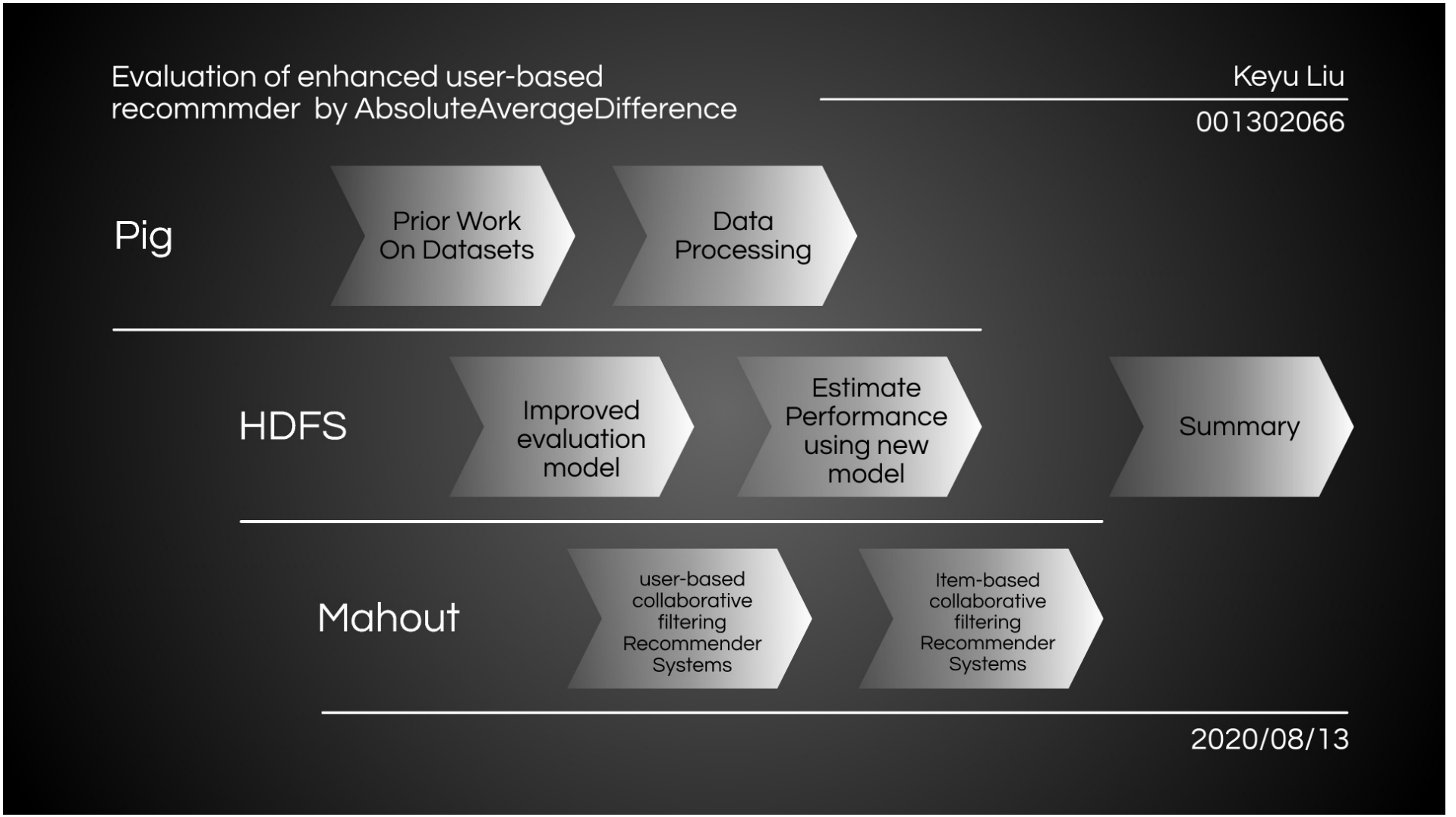


Map Reduce

Discrete Function of AAD and personality

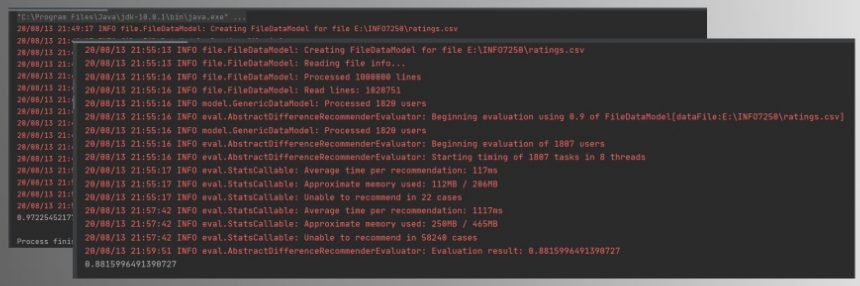
openness	agreeableness	emotional_stability	conscientiousness	extraversion		counter	absoluteDifference	averageAbsoluteDifference	1070203
20	30	40	60	15		1537	15370	10	
35	35	25	65	15		84	840	10	
45	35	50	40	15		1007	10070	10	
50	35	45	30	25		688	6880	10	
55	40	65	25	35		242	2420	10	
55	50	30	50	40		67	670	10	
55	50	65	55	55		1210	12100	10	
55	55	45	35	30		591	5910	10	
55	70	50	60	10		357	3570	10	
65	45	60	40	60		488	4880	10	

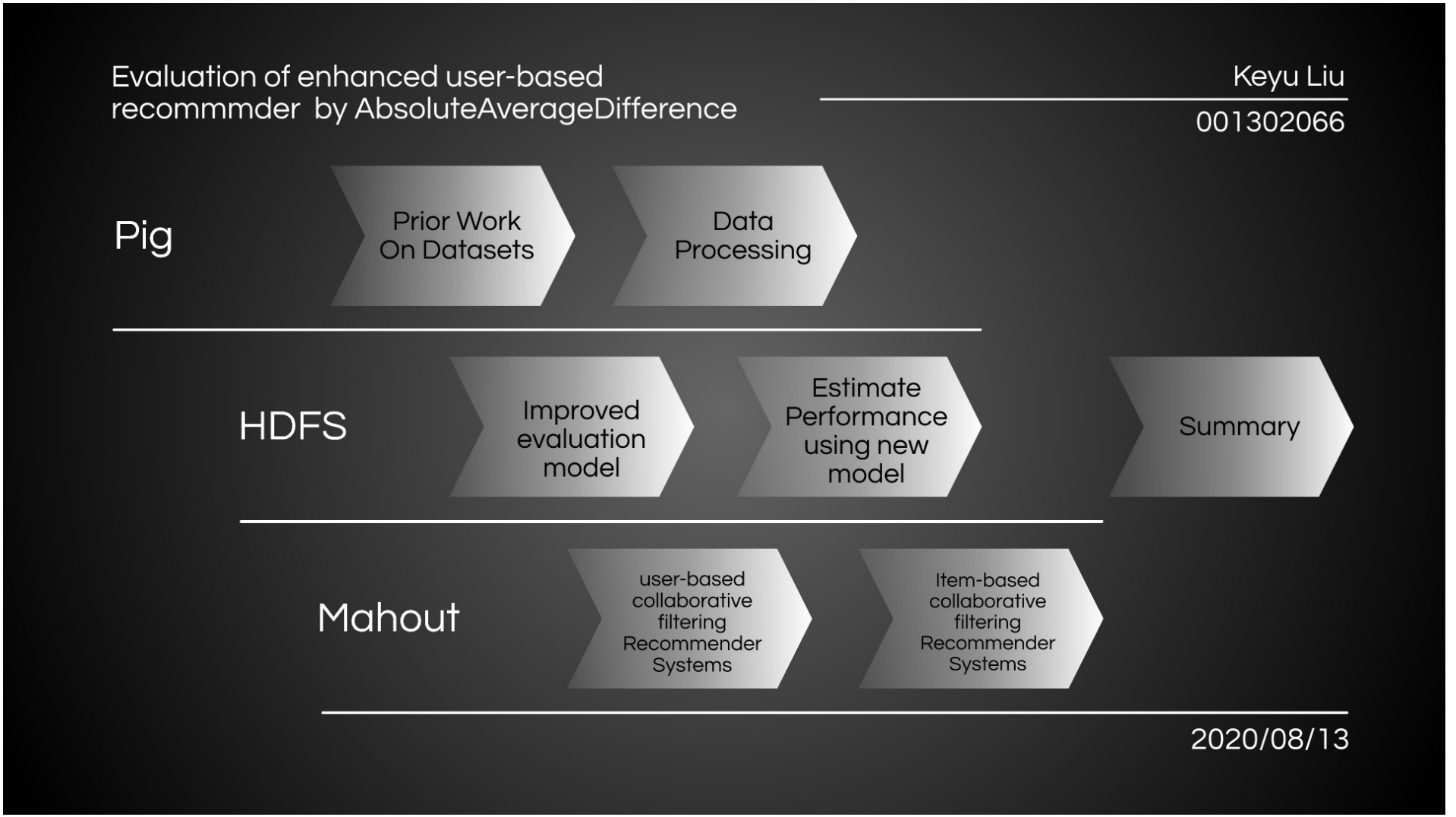
Converting to PDF using distributed linear algebra or knn/SVM. Then, building a model to analysis the relationship between personality and AAD.


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

AbsoluteAverageDifference from hold on test(0.9,0.1)

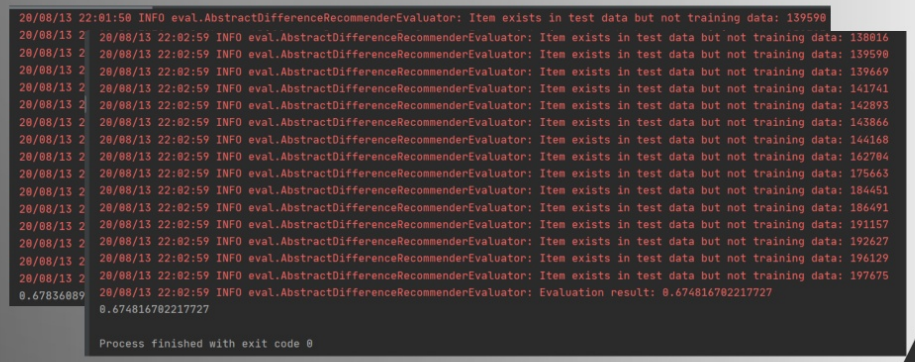
Outout is unstable. But as more test is proceed, the AAD tends to 0.93



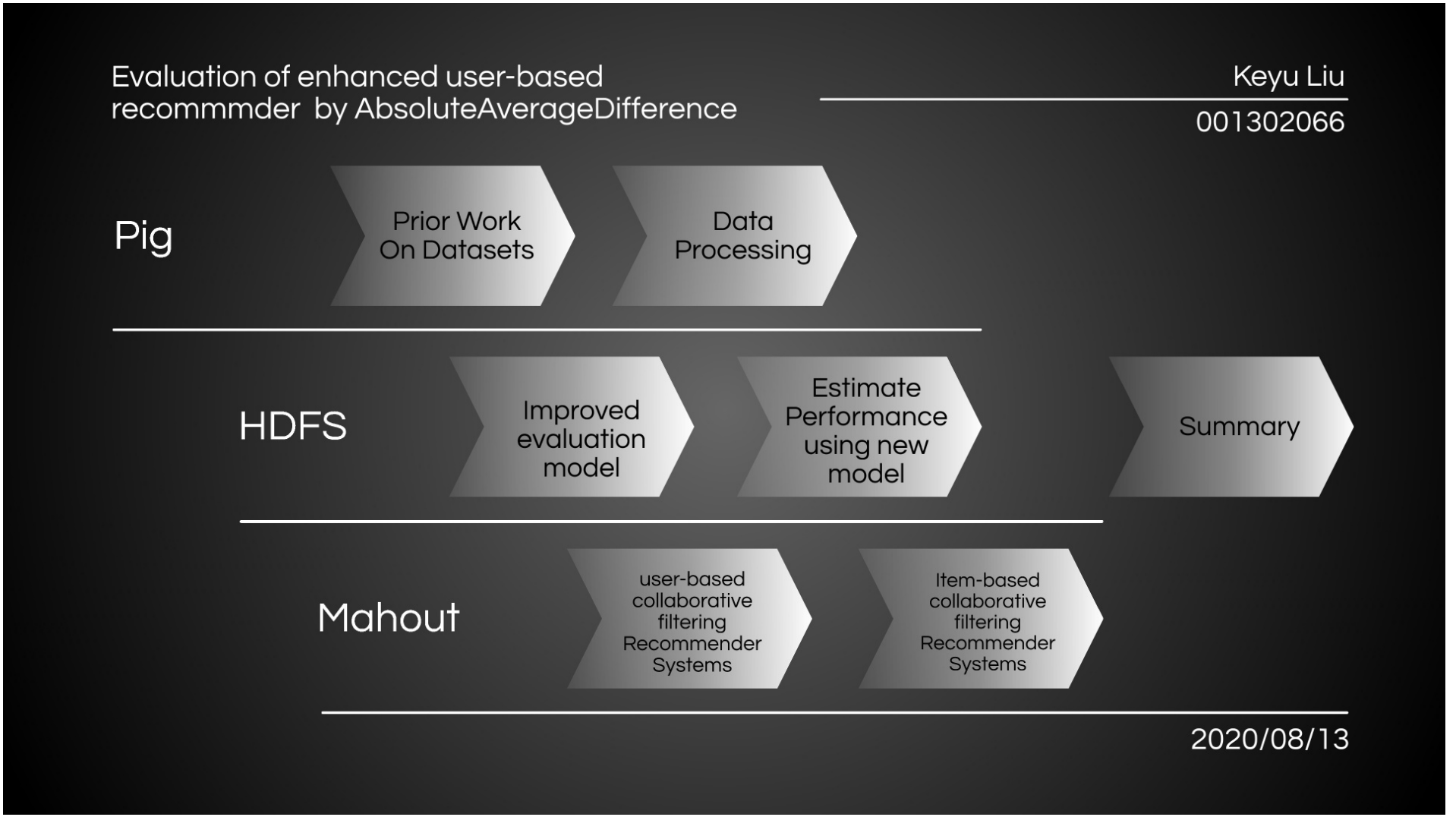

Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

Item-based

Using item itself as if using UserNeighborhood



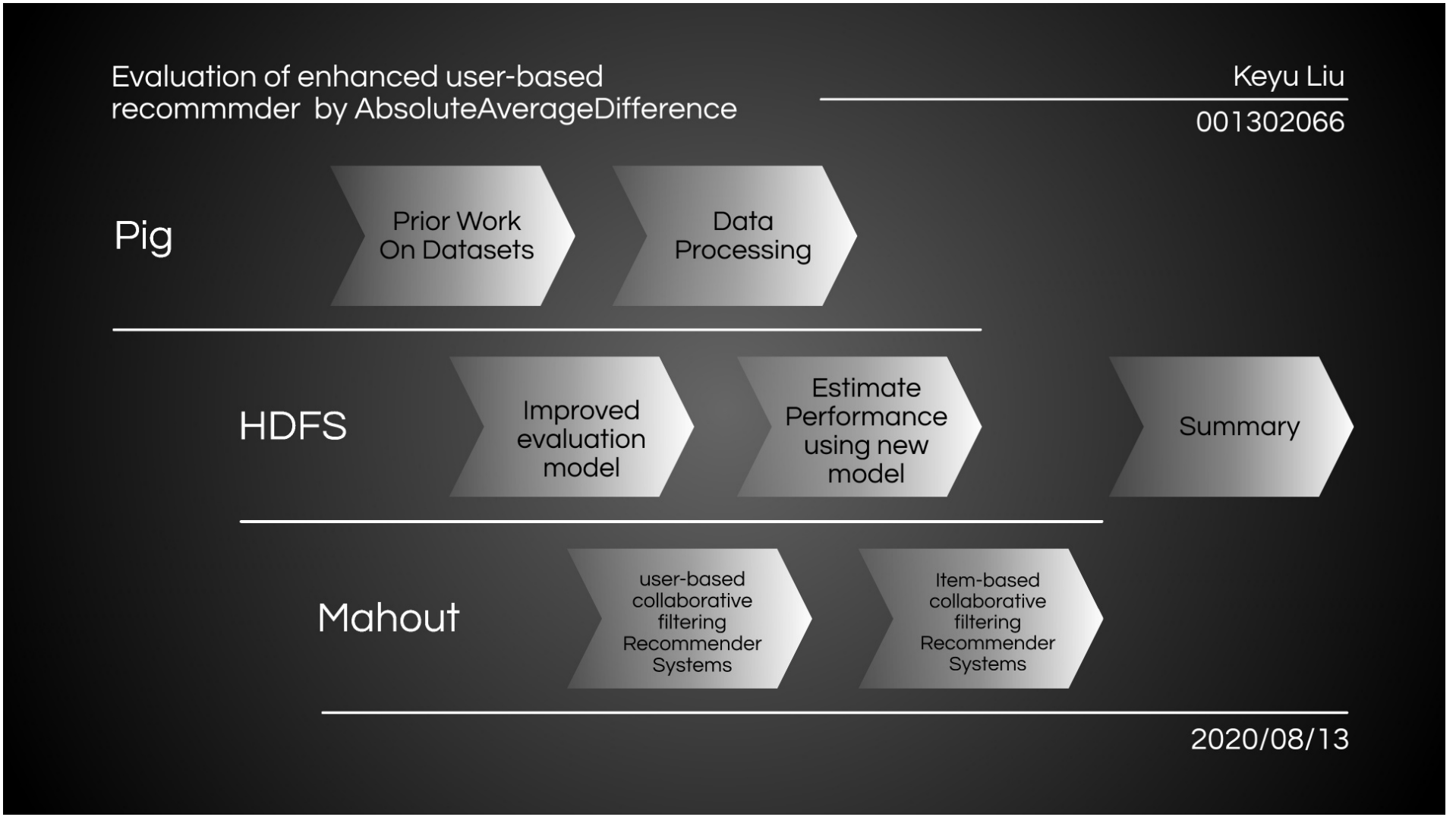
Why output is stable and the prediction is precise ?


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13

Summary

Limitation : can't use predicted movieId with it predicted rating.

Future work: using KNN-item based build new recommender system instead of GenericItemBasedRecommender in API.


Evaluation of enhanced user-based recommmder by AbsoluteAverageDifference
Keyu Liu
001302066
Pig
Prior Work On Datasets
Data Processing
HDFS
Improved evaluation model
Estimate Performance using new model
Summary
Mahout
user-based collaborative filtering Recommender Systems
Item-based collaborative filtering Recommender Systems
2020/08/13