**Results:**

Threshold:0

F1=0.9789

Threshold:0.2

F1=0.9405

Threshold:0.4

F1=0.506

Threshold :0.5

F1=0.136

**Error Analysis:**

*True Positive:*

X result from Y Y complicate by X 0.14017291365999876

X result from Y Y manifest as X 0.1690536975424202

X result from Y Y present as X 0.22971415439983164

X prescribe for Y Y achieve with X 0.1891753833581335

X prescribe for Y X provide from Y 0.29925902949915795

*False Positive:*

X result from Y X produce by Y 0.5213612061380867

X result from Y X cause by Y 0.6270707881354493

X need for Y X require for Y 0.5346499773756952

X observe in Y X occur in Y 0.5176185135711261

X occur with Y Y cause X 0.5188988621781038

*True Negative:*

X resemble Y Y differentiate from X 0.12698443758862624

X resemble Y Y distinguish from X 0.2216033222371107

X simulate Y Y confound with X 0.16572842742137478

X simulate Y Y distinguish from X 0.23022853140962443

X get Y X die from Y 0.24074318112022097

*False Negative:*

X expose to Y X protect from Y 0.5612398309821658

X have Y X be in Y 0.5052111601841854

X produce Y X prevent Y 0.4837122860295577

X protect from Y X expose to Y 0.5612398309821658

X resemble Y X differ from Y 0.4972340512574016

**memory and amount of key value estimation**

**calculating Sim on the negative test set**

step1:

File Input Format Counters

Bytes Read=3514876736

File Output Format Counters

Bytes Written=240647

Map input records=81825916

Map output records=2949720

Map output bytes=211275964

Map output materialized bytes=86498449

Input split bytes=8550

Combine input records=0

Combine output records=0

Reduce input groups=2521039

Reduce shuffle bytes=86498449

Reduce input records=2949720

Reduce output records=6930

Spilled Records=5899440

Shuffled Maps =2625

Failed Shuffles=0

Merged Map outputs=2625

GC time elapsed (ms)=78934

CPU time spent (ms)=1918720

Physical memory (bytes) snapshot=69201534976

Virtual memory (bytes) snapshot=412704636928

Total committed heap usage (bytes)=67379396608

step 2:

File Input Format Counters

Bytes Read=240647

File Output Format Counters

Bytes Written=9923

Map input records=6930

Map output records=6930

Map output bytes=240647

Map output materialized bytes=223047

Input split bytes=4725

Combine input records=6930

Combine output records=6930

Reduce input groups=198

Reduce shuffle bytes=223047

Reduce input records=6930

Reduce output records=198

Spilled Records=13860

Shuffled Maps =1225

Failed Shuffles=0

Merged Map outputs=1225

GC time elapsed (ms)=13705

CPU time spent (ms)=70090

Physical memory (bytes) snapshot=23560445952

Virtual memory (bytes) snapshot=277472022528

Total committed heap usage (bytes)=22687514624

Step 3:

File Input Format Counters

Bytes Read=9923

File Output Format Counters

Bytes Written=4714

Map input records=198

Map output records=198

Map output bytes=9527

Map output materialized bytes=9237

Input split bytes=4760

Combine input records=198

Combine output records=198

Reduce input groups=99

Reduce shuffle bytes=9237

Reduce input records=198

Reduce output records=99

Spilled Records=396

Shuffled Maps =35

Failed Shuffles=0

Merged Map outputs=35

GC time elapsed (ms)=20492

CPU time spent (ms)=47960

Physical memory (bytes) snapshot=20630269952

Virtual memory (bytes) snapshot=121584340992

Total committed heap usage (bytes)=20542128128

**memory and amount of key value estimation**

**calculating Sim on the positve test set**

step1:

File Input Format Counters

Bytes Read=3514876736

File Output Format Counters

Bytes Written=5916121

Map input records=81825916

Map output records=54819335

Map output bytes=3924316251

Map output materialized bytes=1590476355

Input split bytes=8550

Combine input records=0

Combine output records=0

Reduce input groups=48124638

Reduce shuffle bytes=1590476355

Reduce input records=54819335

Reduce output records=173670

Spilled Records=109638670

Shuffled Maps =2625

Failed Shuffles=0

Merged Map outputs=2625

GC time elapsed (ms)=108307

CPU time spent (ms)=3459120

Physical memory (bytes) snapshot=99650146304

Virtual memory (bytes) snapshot=412866220032

Total committed heap usage (bytes)=99308535808

Step2: File Input Format Counters

Bytes Read=5916121

File Output Format Counters

Bytes Written=245397

Map input records=173670

Map output records=173670

Map output bytes=5916121

Map output materialized bytes=2713999

Input split bytes=4690

Combine input records=173670

Combine output records=173670

Reduce input groups=4962

Reduce shuffle bytes=2713999

Reduce input records=173670

Reduce output records=4962

Spilled Records=347340

Shuffled Maps =1225

Failed Shuffles=0

Merged Map outputs=1225

GC time elapsed (ms)=12970

CPU time spent (ms)=112700

Physical memory (bytes) snapshot=23658848256

Virtual memory (bytes) snapshot=277396017152

Total committed heap usage (bytes)=23536861184

Step3:

File Input Format Counters

Bytes Read=245397

File Output Format Counters

Bytes Written=116646

Map input records=4962

Map output records=4962

Map output bytes=235473

Map output materialized bytes=158091

Input split bytes=4725

Combine input records=4962

Combine output records=4962

Reduce input groups=2481

Reduce shuffle bytes=158091

Reduce input records=4962

Reduce output records=2481

Spilled Records=9924

Shuffled Maps =35

Failed Shuffles=0

Merged Map outputs=35

GC time elapsed (ms)=22020

CPU time spent (ms)=52990

Physical memory (bytes) snapshot=21345034240

Virtual memory (bytes) snapshot=121457930240

Total committed heap usage (bytes)=20782252032

**memory and amount of key value estimation on the calculating mi**

File Input Format Counters

Bytes Read=8588063557

File Output Format Counters

Bytes Written=2183392720

Map input records=2025276200

Map output records=3754947632

Combine input records=3754947632

Combine output records=1063736908

Reduce input records=1063736908

Reduce output records=81825916

Physical memory (bytes) snapshot=1369990184960

Virtual memory (bytes) snapshot=4885699379200

Total committed heap usage (bytes)=1366467543040

Step 2:

File Input Format Counters

Bytes Read=2806501788

File Output Format Counters

Bytes Written=2314876731

Map input records=81825916

Map input records=81825916

Map output records=163651832

Map output bytes=5328854250

Map output materialized bytes=1883579289

Input split bytes=10125

Combine input records=0

Combine output records=0

Reduce input groups=86391760

Reduce shuffle bytes=1883579289

Reduce input records=163651832

Reduce output records=81825916

Spilled Records=327303664

Shuffled Maps =5625

Failed Shuffles=0

Merged Map outputs=5625

GC time elapsed (ms)=76704

CPU time spent (ms)=8518470

Physical memory (bytes) snapshot=123444490240

Virtual memory (bytes) snapshot=577316343808

Total committed heap usage (bytes)=117657042944

Step 3:

File Input Format Counters

Bytes Read=2806501788

File Output Format Counters

Bytes Written=3514876736

Map input records=81825916

Map output records=163651832

Map output bytes=5060523198

Map output materialized bytes=2414206857

Input split bytes=10425

Combine input records=0

Combine output records=0

Reduce input groups=82881090

Reduce shuffle bytes=2414206857

Reduce input records=163651832

Reduce output records=81825916

Spilled Records=327303664

Shuffled Maps =5625

Failed Shuffles=0

Merged Map outputs=5625

GC time elapsed (ms)=121020

CPU time spent (ms)=10208940

Physical memory (bytes) snapshot=138450960384

Virtual memory (bytes) snapshot=582866239488

Total committed heap usage (bytes)=122170114048