Recommender Systems – Yelp Kaggle Dataset

yelp_review.csv
(5,261,668)

yelp_business.csv
(174,567)

By Business_Id

reviews_bizname
(5,261,668)

Filter:
1. State = 'NV'
2. Category like '%Restaurant%'
3. Sample 2000 unique restaurants

reviews_bizname2
(289,418)

NLP on Review Text
for Verbs and Nouns

res
(2,000)

TI-DIF Dtm

Document Matrix
(2,000 x 12,192)

Preferences...

Keywords NLP

**Keywords
Dataframe**

Calculate 'Cosine'
distance between
keywords & TIDIF of
Review Data

Top 500 Restaurants

**Choice
Universe**

Cuisines..

Filter :
Category like '%
<Cuisinine>%'

**Serendipity Store for 5 restaurants
reviews_bizname2 - resSubset2.1**
(1,756)

resSubset2
(8,102)

<=30

resSubset3
(9,858)

>30

Count of Unique
Restaurnts

<= 30

Filter :
Restaurant in
'%ChoiceUniverse%'

resSubset2.1
(7,445)

Inputs

User

1. Find 3 most reviewed users
2. Find associated categories
3. DTM - Wieghted bin
4. Exclude - Input Cuisine and Res / Food
5. Score based on the ratings for each
6. Recommend top 10 restaurants in descending score

CB Model

1.Build Rating Matrix
2.Recommend top 10 restaurants from
the closest match

CF Model

Output /
Recommendations

Top 10 recommendation
based on normalized
score of RMSE results
from both.