Search, Relevance, and Recommendation for Wine Reviews

CS-410 Fall 23'
qinxiw2@illinois.edu

Roadmap

1. Motivation
2. Data Preprocessing
3. Sentiment Analysis
4. Recommender System
5. Search and Retrieval
6. Topic Identification
7. Software Usage Tutorial
8. Discussion

1. Motivation

Combine and utilize various things we have learned this semester, including search, retrieval and ranking, NLP techniques, text classification/categorization, and recommendations:

1) Search wine by description using free text;

2) Categorize reviews by positive, neutral, or critica sentiment;

3) Retrieve relevant wine search by variety, country, designation, etc;

4) Recommend similar wines matched via content based on review and attribute similarities

Prerequisite on installation

git clone https://github.com/QinxiW/TISProject.git

Then from the root dir....

- Create a python 3.11 virtualenv

  python3 -m venv myenv

- Activate the virtualenv

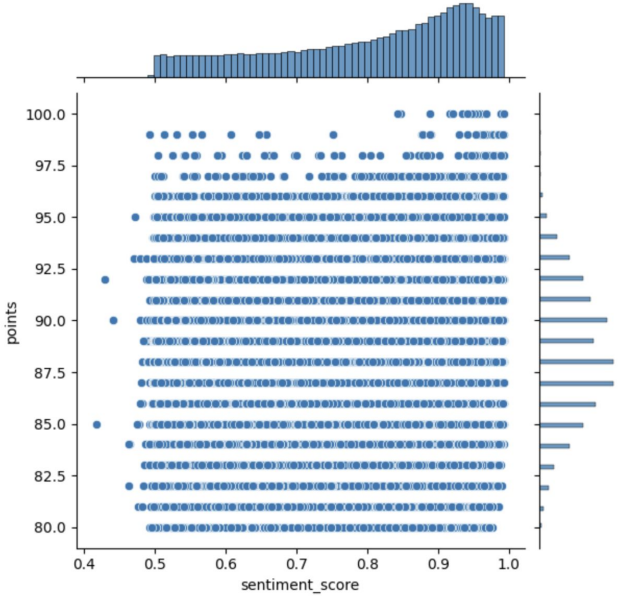
  source myenv/bin/activate

- Install all the necessary libraries needed

  pip install -r requirements.txt

2. Data Preprocessing

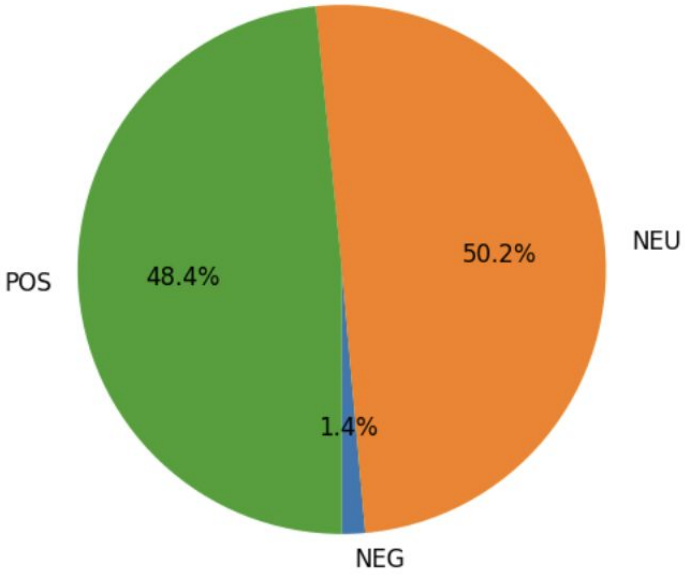
WineReviews	
#	bigserial
country	text
description	text
designation	text
points	bigint
price	float
province	text
region_1	text
region_2	text
taster_name	text
taster_twitter_handle	text
title	text
variety	text
winery	text



*python Data/data_cleaning.py -- file_location_path {dataset}**

**due to the size of the original dataset, we did not check in the file into the source code. You can download it from https://www.kaggle.com/datasets/zynicide/wine-reviews/data. This is not a requirement, and you do not need this raw dataset to run any of the later scripts, as the output data is checked in and baked into the code already*

3. Sentiment Analysis



python Model/sentiment_analysis_and_fine_tuning.py

```
Accuracy: 0.5858994124755198
Confusion Matrix:
 [[   3  212  131]
 [  10 8050 3961]
 [   7 5617 6008]]
Classification Report:
              precision    recall  f1-score   support

         NEG       0.15      0.01      0.02       346
         NEU       0.58      0.67      0.62     12021
         POS       0.59      0.52      0.55     11632

    accuracy                           0.59     23999
   macro avg       0.44      0.40      0.40     23999
weighted avg       0.58      0.59      0.58     23999
```

4. Recommender System - Collaborative Filtering

python Model/cf_recommender.py

```
recommend(item_id=itemId_, num=5)
```

```
Using itemId 82886 which is Pascal Bouchard 2007 Les Clos Grand Cru  (Chablis)

Recommending 5 products similar to Pascal Bouchard 2007 Les Clos Grand Cru  (Chablis) ...
-----
Recommended: Tegernseerhof 2006 Steinertal Riesling (Wachau)(score: 0.11)
Recommended: Château la Varière 2009 Les Melleresses  (Bonnezeaux)(score: 0.09)
Recommended: Steininger 2007 Grand Grü Grüner Veltliner (Kamptal)(score: 0.09)
Recommended: Domaine Laroche 2009 Les Vaudevey  (Chablis)(score: 0.08)
Recommended: Domaine Pascal et Mireille Renaud 2015 Aux Insarts  (Pouilly-Fuissé)(score: 0.08)
```

```
First wine description:
A super-rich wine, packed with peach and green plum fruits, layers of wood and an intense concentration. An impressive wine that shows weight and richnes

Matched wine description:
The intensity of this wine is almost overwhelming, with its crisp minerality, apple skin and green plum flavors and white stone fruits. The concentration
```

4. Recommender System - Random Forest

python Model/rf_recommender.py

```
Loading data from /Data dir
df loaded, head:     Unnamed: 0.1  Unnamed: 0  ...                          description_cleaned_limit                          base_sentiment_label
0             0           0  ...  aromas include tropical fruit  broom  brimston...  {'label': 'NEU', 'score': 0.8123481869697571}
1             1           1  ...  this is ripe and fruity  a wine that is smooth...  {'label': 'POS', 'score': 0.7544678449630737}
2             2           2  ...  tart and snappy  the flavors of lime flesh and...  {'label': 'NEU', 'score': 0.9003725647926331}
3             3           3  ...  pineapple rind  lemon pith and orange blossom ...  {'label': 'NEU', 'score': 0.6538661122322083}
4             4           4  ...  much like the regular bottling from 2012  this...  {'label': 'NEU', 'score': 0.9240036606788635}

[5 rows x 41 columns]
Accuracy: 0.9043293470561273
Precision: 0.5382775119617225
Recall: 0.4580618892508143
Confusion Matrix:
 [[20578   965]
 [ 1331  1125]]
Classification Report:
              precision    recall  f1-score   support

       False       0.94      0.96      0.95     21543
        True       0.54      0.46      0.49      2456

    accuracy                           0.90     23999
   macro avg       0.74      0.71      0.72     23999
weighted avg       0.90      0.90      0.90     23999
```

4. Recommender System - k-nearest neighbors

```
Recmmendation for Mandilaria:

1: Plyto with distance: 0.2810116239508881
2: Kotsifali with distance: 0.2810116239508881
3: Thrapsathiri with distance: 0.2810116239508881
4: Vilana with distance: 0.2810116239508881
5: Assyrtiko with distance: 0.6445111635507379


Recmmendation for Pinot Blanc-Chardonnay:

1: Muskat with distance: 0.18098001617005233
2: Edelzwicker with distance: 0.3473356334124358
3: Blauburgunder with distance: 0.4073943606866912
4: Tokay Pinot Gris with distance: 0.42714788199713494
5: Alsace white blend with distance: 0.42714788199713494


Recmmendation for Cabernet Sauvignon-Carmenère:

1: Syrah-Cabernet with distance: 0.3727085611068949
2: Carmenère with distance: 0.408495303989592
3: Carmenère-Cabernet Sauvignon with distance: 0.4149521276079837
4: Carmenère-Syrah with distance: 0.5319680561775886
5: Cabernet Blend with distance: 0.5877003569675064
```

python Model/knn_recommender.py

5. Search & Retrieval

python Search/wine_search.py

```
Welcome to the wine community! We hope you will find wine you enjoy here via search and recommendation.
Let's start by telling me about what wine you are looking for. Type exit to leave anytime.
you can search with a short phrases what type of wine (e.g 'sherry', 'italian') or wine attributes (e.g
 'aromatic', 'fruit flavors'): sherry

Got it! Based on  sherry  we found the following wine title based on your search:

1 :  Pedro Romero NV Full, Dry Oloroso Sherry (Jerez) from Spain- $13
2 :  Domaine de Savagny 2011 Savagnin (Côtes du Jura) from France- $52
3 :  González Byass NV Solera 1847 Cream Sherry (Jerez) from Spain- $26
```

6. Topic Identification & Suggestion

python Search/wine_inference.py

```
For other wines similar to your search, we also recommend:

1 :  Osborne NV Cream Sherry (Jerez) from Spain- $18
2 :  Delgado Zuleta NV Premium Fino Sherry (Jerez) from Spain- $16
3 :  Lustau NV Dry Amontillado Los Arcos Sherry (Jerez) from Spain- $17


Here are some wine variety we think you will like: palomino, with recommendation score: 75%
We found these top common words for the matched wine reviews, if you find anything you like you can
search more with them:
'brown sugar, stone fruits, dried apricot, caramel toffee, nutty flavors, palate flavors', 'garnacha
 blanca, fruit flavors, acidity flavors, white garnacha, peach pit, green apple', 'flavors lemon, le
mon green, apple peach, flavors finish, tropical fruit'


Any other wine search? Type exit to leave anytime:
```

7. Software Usage Tutorial - Demo

8. Discussion

- Accomplishments
  - We built an interactive program that allows search and other exploratory activities with the wine and their attributes(variety, region, etc) and reviews.
- Observations
  - Lack labeled data for accurate metrics in evaluation, hard to compare different models
- Next steps
  - Gather user feedback and add instrumentation of online metrics
  - Use user feedback to run transfer learning