

Wine Recommendation System

BIG DATA AND ANALYTICS

B.E. COMPUTER ENGINEERING

Guide:

Ms. Anuradha Srinivasaraghavan

By

Kinen Cardoza 09

Velina Cutinha 14

Arnold Dsouza 23

Misaal D'souza 25



Department of Information Technology

St. Francis Institute of Technology

(Engineering College)

University of Mumbai

Contents

Chapter	Contents	Page No.
1	PROBLEM STATEMENT	3
2	OBJECTIVE	3
3	EXPERIMENTAL SETUP	3-4
4	IMPLEMENTATION	4-8
5	VISUALIZATION	9-10
6	RESULT AND ANALYSIS	11
7	CONCLUSION	11

01. Problem Statement:

To develop a wine recommendation system that will recommend wine based on variety. The user will input one type of variety of a wine and the system will recommend from 756 different types of variety.

02. Objective:

- Our main objective is to recommend user variety of wine based on the preference of the user.
- To find the most and least expensive wine by country.
- To find the count of wine tasters from different countries.
- To look for common words used to describe wine.
- To present reviews of top wine tasters based on wine varieties.

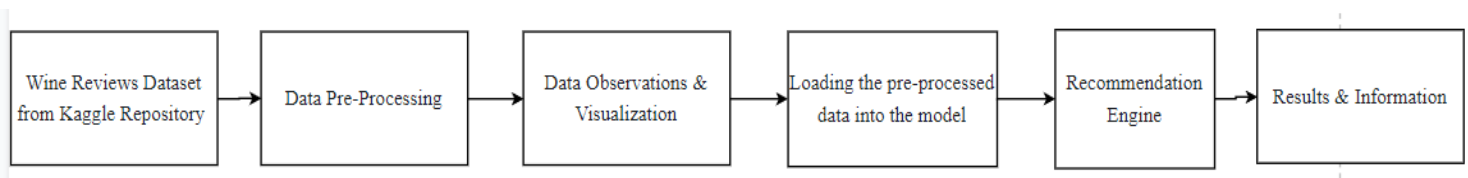


Fig1: Proposed Architecture

03. Experimental Setup:

- Software Requirements-
 - Anaconda or Command Prompt
 - Colab or Jupyter Notebooks
- Hardware Requirements-
 - Hard Disk- 256 GB or more
 - RAM - 4 GB or above
 - Processor - Core i3 or above
- Technologies Used -
 - Python
 - Recommendation Engine

Dataset Description:

Data Selection: The required data set is collected from Kaggle Databases.

The dataset has 280901 records. It consists of 13 features.

Features :

country, description, designation, points, price, province, region_1, region_2, variety, winery, taster_name, taster_twitter_handle, title.

country	description	designation	points	price	province	region_1	region_2	variety	winery	taster_name	taster_twitter_handle	title
US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz	NaN	NaN	NaN
Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro	NaN	Tinta de Toro	Bodega Carmen Rodríguez	NaN	NaN	NaN
US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley	NaN	NaN	NaN
US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi	NaN	NaN	NaN
France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol	NaN	Provence red blend	Domaine de la Bégude	NaN	NaN	NaN
...
Germany	Notes of honeysuckle and cantaloupe sweeten th...	Brauneberger Juffer-Sonnenuhr Spätlese	90	28.0	Mosel	NaN	NaN	Riesling	Dr. H. Thanisch (Erben Müller-Burggraef)	Anna Lee C. Iijima	NaN	Dr. H. Thanisch (Erben Müller-Burggraef) 2013 ...
US	Citation is given as much as a decade of bottl...	NaN	90	75.0	Oregon	Oregon	Oregon Other	Pinot Noir	Citation	Paul Gregutt	@paulgwine	Citation 2004 Pinot Noir (Oregon)
France	Well-drained gravel soil gives this wine its c...	Kritt	90	30.0	Alsace	Alsace	NaN	Gewürztraminer	Domaine Gresser	Roger Voss	@vossroger	Domaine Gresser 2013 Kritt Gewurztraminer (Als...
France	A dry style of Pinot Gris, this is crisp with ...	NaN	90	32.0	Alsace	Alsace	NaN	Pinot Gris	Domaine Marcel Deiss	Roger Voss	@vossroger	Domaine Marcel Deiss 2012 Pinot Gris (Alsace)
France	Big, rich and off-dry, this is powered by inte...	Lieu-dit Harth Cuvée Caroline	90	21.0	Alsace	Alsace	NaN	Gewürztraminer	Domaine Schoffit	Roger Voss	@vossroger	Domaine Schoffit 2012 Lieu-dit Harth Cuvée Car...

04. Implementation Code:

Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import pickle
import streamlit as st
import cssutils
from sklearn.neighbors import NearestNeighbors
from scipy.sparse import csr_matrix
from sklearn.decomposition import TruncatedSVD
```

```
from PIL import Image

# Path of Application Folder

path = "C:/Users/misaa/OneDrive/Desktop/WineRecommendationSystem/"

df = pd.read_csv(path + "wine.csv")

col = ['province', 'variety', 'points']

wine1 = df[col]

wine1 = wine1.dropna(axis=0)

wine1 = wine1.drop_duplicates(['province', 'variety'])

wine1 = wine1[wine1['points'] > 85]

wine_pivot = wine1.pivot(index= 'variety', columns='province', values='points').fillna(0)

wine_pivot_matrix = csr_matrix(wine_pivot)

wine_pivot_find = wine_pivot.copy()

wine_pivot_find.reset_index(level=0, inplace=True)

# Model

knn = NearestNeighbors(n_neighbors=10, algorithm= 'brute', metric= 'cosine')

model_knn = knn.fit(wine_pivot_matrix)

# Main Function

results = []

def wine_recommendation(var):

    query_index = wine_pivot_find.index[wine_pivot_find.variety == var]

    arr = query_index.values

    for i in range(1):

        query_index = arr[i]

        distance, indice = model_knn.kneighbors(wine_pivot.iloc[query_index,:].values.reshape(1,-1), n_neighbors=6)

        for i in range(0, len(distance.flatten())):
```

```

    if i!=0:

        x = i,wine_pivot.index[indice.flatten()[i]]

        results.append(x[1])

    return results

def predict_wrs(var):

    one = 'https://www.google.com/search?q='

    r = wine_recommendation(var)

    st.subheader("Recommendations for "+var+":\n")

    for i in range(0,5):

        x = r[i]

        y = x # With spaces

        if ' ' in y == True :

            y = y.replace(' ','+')

        link = one + y.replace(' ','+')

        st.write('* [{}]({})'.format(y,link))

# Pickle : write and read

pickle_out = open(path+"predict_wrs.pkl", "wb")

pickle.dump(predict_wrs, pickle_out)

pickle_out.close()

pickle_in = open(path+'predict_wrs.pkl', 'rb')

classifier = pickle.load(pickle_in)

# Input

st.title('Wine Recommendation System')

st.subheader('Variety Name:')

wine_name = st.text_input("")

```

```
submit = st.button('Predict')

if submit:
    predict_wrs(wine_name)

# set index to empty strings
df.index = [""] * len(df)

st.table(df)

# Bordeaux

st.write("* Bordeaux-style Red Blend")

col1, mid, col2 = st.beta_columns([3,1,14])

with col1:
    st.image('images/bordeaux.jpg', width=125)

with col2:
    st.write("Bordeaux type of grapes are used for the most expensive wines.")

# Portuguese Red

st.write("* Portuguese Red")

col1, mid, col2 = st.beta_columns([3,1,14])

with col1:
    st.image('images/portuguese.jpg', width=125)

with col2:
    st.write("Portuguese grapes are found to be used in the wines that have the highest ratings from the wine tasters.")

# image

st.subheader('Visualization')

# viz 01

st.write('* Most & Least Expensive Wine by Country')

img = Image.open("images/most_exp.png")
```

```
st.image(img)
```

```
# viz 02
```

```
st.write('* Most Expensive & Most Rated Wine Prepared by Country')
```

```
img = Image.open("images/most_prep.png")
```

```
st.image(img)
```

```
# viz 03
```

```
st.write('* Wine Taster Count by Country')
```

```
img = Image.open("images/taster_country.png")
```

```
st.image(img)
```

```
# viz 04
```

```
st.write('* Common Words Used to Describe Wine')
```

```
img = Image.open("images/common_words.png")
```


```
st.image(img)
```


05. Vizualization:

- Frontend :

Observations

- France produces nearly every strain of international grape and it also harbors the most [expensive wine](#) in the world.



- US is the highest producer of wine in the world with a count of 54,504 followed by France with 22,093
- [Cabernet Sauvignon](#), [Pinot Noir](#), [Airen](#), [Merlot](#) are the most widely used grape varieties in the world.
- The countries that produce the

Wine Recommendation System

Variety Name:

Apple

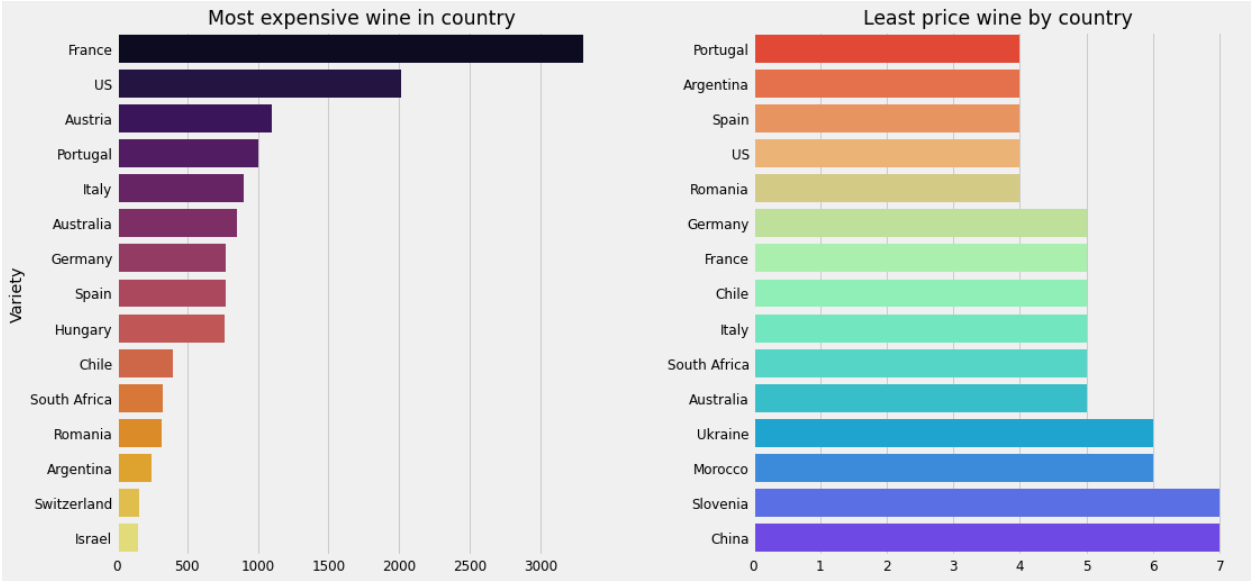
Predict

Recommendations for Apple:

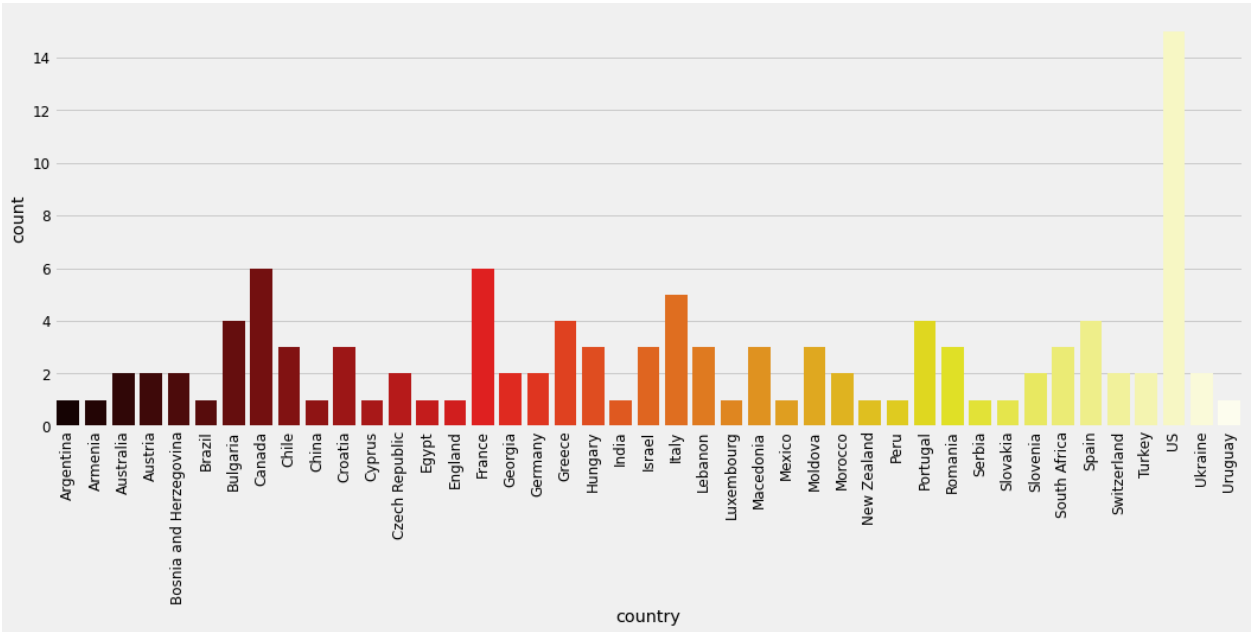
- [Refosco](#)
- [Rhône-style White Blend](#)
- [Rhône-style Red Blend](#)
- [Ribolla Gialla](#)
- [Red Blend](#)

Information

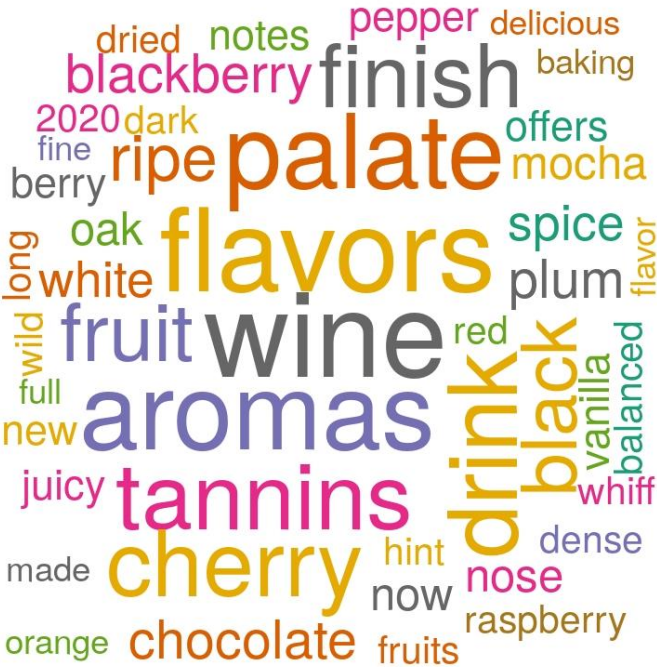
- Most expensive and least expensive of wine by country



Wine Taster Count By Country



Common Words Used to Describe Wine



06. Results and Analysis:

We used the wine reviews dataset from the Kaggle repository and made a recommendation system. The user is able to enter a variety of wine and the system predicts the result. The user can also get more information of different wine varieties by simply clicking one of the recommended options. We also made several analysis such as ranking of main grape varieties planted worldwide, most and least expensive wine by country, most expensive and most rated wine prepared by country, wine taster count by country, common words used to describe wine and reviews from top wine tasters based on varieties.

07. Conclusion:

The recommendation engine uses a collaborative filtering method to predict what the user will like. We successfully developed a wine recommendation system that gives user preferences based on the wine variety given by the user. Several analysis were made based on different features from the dataset such as *country*, *taster_name*, *description*, *winery*, etc. Hence, we successfully developed an application that recommends users wine varieties and even made various analysis on the same.