Georgiy Voropayev T00570108

COMP 3710 Artificial Intelligence

Personal Project

November 11, 2019

Proposal of My Personal Project

Wine Classification

As the wine became more and more popular through the ages, more and more companies from all around the world have started the production of their own wines with the subtlety and delicacy specific to that area. Various governments started to implement laws and guidance’s that regulate their classification. Back in the ancient times, Roman winemakers started cataloging all the possible methods for creation and uses of wine, which was greatly improved when winemaking spread beyond Europe. By then, Italian and French governments started implementing strict laws, which helped to authenticate traditional wines that were made in specific areas of their countries, or by using specific grapes and techniques. The ultimate classification categories that we regularized are **place of origin** (or appellation)**, vinification** method and style, **sweetness, taste, quality,** and **vintage** or **varietal** (which describes from what variety of grapes was selected wine been made).

**Datasets:**

I chose “Wine Reviews” dataset, which contains two CSV files, winemag-data-130k-v2.csv and winemag-data\_first 150k.csv. The data of all the wines is displayed in columns such as **country, description, designation, points, price, province, region\_1, region\_2, taster\_name, taster\_twitter\_handle, title, variety,** and **winery.** This data was gathered from WineEnthusiast website.

**Reason and Goal:**

The reason I chose these datasets is due to my interest in wines, and what contributes to their tastes and classification. In my opinion, it is practical and helpful to have a trained program that can classify a set of wines it has never seen before. My goal is to create a program that takes criteria such as description, place of origin, etc. (listed above) and would predict the variety of a specific wine.

**Technologies and Packages:**

For my project I will be using Scikit-learn along with pandas. Pandas is a software library for python programming language that is used for data manipulation and analysis. This library is useful to my project development not only because I will be dealing with data manipulation, but also because my datasets contain almost no numbers, but words.