PREDICTIONS OF WINE RATINGS USING MACHINE LEARNING ALGORITHMS

Mai Bui, Albers School of Business, Seattle University, 901 12th Avenue, Seattle, WA 98122,
mbui@seattleu.edu
Bryce Dorrance, Albers School of Business, Seattle University, 901 12th Avenue, Seattle, WA 98122,
dorrance@seattleu.edu
Sofia Rodriguez, Albers School of Business, Seattle University, 901 12th Avenue, Seattle, WA 98122,
srodriguez@seattleu.edu
Marc Winter, Albers School of Business, Seattle University, 901 12th Avenue, Seattle, WA 98122
mwinter@seattleu.edu
Ben Kim, Albers School of Business, Seattle University, 901 12th Avenue, Seattle, WA 98122,
bkim@seattleu.edu

ABSTRACT

The goal of this paper is to build machine learning models to predict the wine rating scores. For this paper, we built the five models using the decision tree, random forest, neural network, gradient boosting, and clustering algorithms to predict the ratings using the data such as geographic locations, price, variety, temperature, longitude, latitude, country, and critics' textual reviews from the multiple datasets. The wine reviews required natural language processing (NLP). For evaluation of the models, the K-fold scores method was used. We found that the gradient boosting algorithm produced the highest performance for predictions. The datasets we used include the wine reviews available from Kaggle, geographical names, global land temperatures, and world cities.

Keywords: wine ratings, machine learning, natural language processing, gradient boosting