Issues in Information Systems
Volume 24, Issue 1, pp. 142-146, 2023

DOI: https://doi.org/10.48009/1_iis_2023_112

Text mining of textual descriptions of scotch whisky ratings

Ben Kim, *Seattle University, bkim@seattleu.edu*

Abstract

This study analyzed the textual descriptions of Scotch Whiskies with their rating scores and descriptions. We applied natural language processing and built three machine learning models using Decision Tree, Gradient Boosting, and Random Forest. Consumers of whiskies use rating scores and descriptions for their purchasing decisions. We were interested in whether prices, ratings, and descriptions provide reliable information for consumers. After calculating the R-squared values from those machine learning models, we found that the prices, rating scores, and descriptions are minimally consistent with each other.

Keywords: whisky ratings, natural language processing, text mining, decision tree, gradient boosting, random forest

Introduction

When consumers purchase Scotchy Whiskies, they check the ratings, descriptions (commentaries of those ratings) as well as prices. One example of those ratings can be found in Table 1. Using the large amount of data available, this study was interested in testing whether these ratings and descriptions are reliable and the prices are consistent with those ratings. Most studies on whiskies used chemical compositions or sensory information. Instead, the descriptions were analyzed by natural language processing (NLP).

Some examples of descriptions are shown in Table 2. For the analysis, the textual descriptions were converted into a matrix of numbers (known as the *Document Term Matrix*) and applied three machine learning (ML) algorithms – Decision Tree, Gradient Boosting, and Random Forest to build predictive models. R-Squared (R^2) scores was used to evaluate these models.

Table 1: Rating Scheme

Review Point	Recommendation
95-100	Classic: a great whisky
90-94	Outstanding: a whisky of superior character and style
85-89	Very good: a whisky with special qualities
80-84	Good: a solid, well-made whisky
75-79	Mediocre: a drinkable whisky that may have minor flaws
50-74	Not recommended

Source: https://www.whiskyadvocate.com/ratings-and-reviews/

Background and literature review

Using the advance in artificial intelligence can help the consumer assessments of food and beverages (Fuentes, 2022). Researchers are now developing an artificial tongue distinguishing between various brands of whisky such as Glenfiddich, Glen Marnoch, and Laphroaig with 99 % accuracy (Mone, 2020). Also, general olfactory perceptual descriptors such as “fishy”, “floral”, or “fruity” can be used for accurate inference of smell of odorous molecules (Gutiérrez et al., 2018). Macias et al. (2019) developed an artificial (bimetallic) tongue to differentiate off the shelf whiskies with > 99.7 % accuracy. Most of the research articles about whiskeys using machine learning algorithms have used non-textual information to analyze and identify the whiskeys.

Table 2: Examples of Whisky Descriptions

"Magnificently powerful and intense. Caramels, dried peats, elegant cigar smoke, seeds scraped from vanilla beans, brand new pencils, peppercorn, coriander seeds, and star anise make for a deeply satisfying nosing experience. Silky caramels, bountiful fruits of ripe peach, stewed apple, orange pith, and pervasive smoke with elements of burnt tobacco. An abiding finish of smoke, dry spices, and banoffee pie sweetness. Close to perfection. Editor's Choice"
"What impresses me most is how this whisky evolves; it's incredibly complex. On the nose and palate, this is a thick, viscous, whisky with notes of sticky toffee, earthy oak, fig cake, roasted nuts, fallen fruit, pancake batter, black cherry, ripe peach, dark chocolate-covered espresso bean, polished leather, tobacco, a hint of wild game, and lingering, leafy damp kiln smoke. Flavors continue on the palate long after swallowing. This is what we all hope for (and dream of) in an older whisky!"
"There have been some legendary Bowmores from the mid-60s and this is every bit their equal. All of them share a remarkable aroma of tropical fruit, which here moves into hallucinatory intensity: guava, mango, peach, pineapple, grapefruit. There’s a very light touch of peat smoke, more a memory of Islay than the reality. Concentrated; even at low strength the palate is silky, heady, and haunting, and lasts forever in the dry glass. A legend is born. (Eight bottles only for the U.S.) Editor's Choice.
"With a name inspired by a 1926 Buster Keaton movie, only 1,698 bottles produced, and the news that one of the two batches is more than 30 years old, the clues were there that this blend was never going to be cheap. It isn't, but it's superb, rich in flavor that screams dusty old oak office, fresh polish, and Sunday church, with spices, oak dried fruits, squiggly raisins, and a surprising melting fruit-and-nut dairy chocolate back story."

Robert Parker, who is one of the most prominent wine critics, stated that "Scores, however, do not reveal the important facts about a wine. The written commentary that accompanies the ratings is a better source of information regarding the wine's style and personality, its relative quality vis-à-vis its peers, and its value and aging potential than any score could ever indicate" (Parker, 2022). The written commentaries were the ones we were interested in for our analysis. Kim (2022) showed the R-squared value between the wine descriptions by sommeliers and rating scores at approximately .6.

This study was interested in how much the descriptions of whiskies can explain their ratings. One of the major differences in ratings between wines and whiskies would be the qualifications of sommeliers. To become a wine sommelier, it requires many years of training and rigorous examinations. However, that is not necessarily the case for whiskey sommeliers (Peters, 2019). In this paper, we examined the consistency of ratings and commentaries (or descriptions) by whiskey sommeliers.

Table 3: Categories of Whiskies

Category	Number of Whiskies
Single Malt Scotch	1819
Blended Scotch Whisky	211
Blended Malt Scotch Whisky	132
Single Grain Whisky	57
Grain Scotch Whisky	28

Dataset and preprocessing of data

This paper used a dataset available at Kaggle (2022). The file name is "*scotch.csv*." It has 2,247 rows and 7 columns. The size of the file is 412 kilobytes. The columns in the file include 'name', 'category', 'review.point', 'price', 'currency', and 'description'. 'name' shows the brand names of whiskies. There are 2223 unique names such as Johnnie Walker Blue Label, Black Bowmore, and others. As for the 'category' column, it has five unique categories as shown in Table 3. The 'review.point' column shows the rating scores given by the reviewers. The maximum point value is 100. The name of this column was to 'review_point' to be consistent with the conventions for Python programming. 'price' is the price of each bottle. The correlation coefficient between price and review point is quite low at 0.12 unlike our initial expectation. The mean, median, standard deviation, maximum, and minimum values of 'review_point' and 'price' are shown in Table 4.

Table 4: Statistics for Review Point and Price

Category	Mean	Median	Maximum	Minimum	Standard Deviation
Review Point	86.7	87	97	63	4.05
Price	643.12	110	157000	12	4702.63

While the columns 'price' and 'review_points' contain numerical data, the 'description' column has text data. These descriptions were used as one of the predictors in our ML models for prices and the review points. Table 4 shows some examples of these descriptions. There are no null values in the dataset. After removing the obviously irrelevant data columns ('unnamed: 0' and 'currency'), the dataset has 2247 rows and 5 columns. To apply ML algorithms, we converted the nominal attribute ('category') to five multiple binary columns (dummy variables) using the *One Hot Encoding* technique.

Natural language processing for whisky descriptions

Since whisky descriptions were written in a natural language (English), we needed to convert them to a matrix of numbers - *Document Term Matrix*. There are at least two methods available in the scikit-learn library for Python – *CountVectorizer* and *TfidfVectorizer*. *TfidfVectorizer* implements the term frequency-inverse document frequency(tf-idf) method. In tf-idf, the words frequently appearing in many documents have lower weights than the less frequent ones. The vectorizer generated 8,818 terms and the document term matrix produced had 2,247 rows and 8,818 columns.

Table 5: Performances for each Algorithm and Set of Features

Predicted ~ Predictors	Algorithms	R-Squared
review_point ~ price	Decision Tree	0.0072
	Gradient Boosting	0.1618
	Random Forest	0.0560
price ~ review_point	Decision Tree	-0.0241
	Gradient Boosting	-0.0237
	Random Forest	-0.0309
review_point ~description	Decision Tree	-0.5206
	Gradient Boosting	0.2126
	Random Forest	0.2582
price ~ description	Decision Tree	-0.6685
	Gradient Boosting	-0.1396
	Random Forest	-0.0222
price ~ description, category	Decision Tree	-0.9031
	Gradient Boosting	-0.2226
	Random Forest	-0.0104
review_point ~description, category	Decision Tree	-0.6069
	Gradient Boosting	0.2106
	Random Forest	0.2565
Review_point ~ escription, category, price	Decision Tree	-0.3343
	Gradient Boosting	0.3276
	Random Forest	0.3047

Data models and discussion

As for the data to apply the ML algorithms, this study used seven datasets for predictors (independent variables) and predicted variable (dependent variable) as shown in Table 5. For these seven datasets, three ML algorithms were applied – Decision Tree, Gradient Boosting, and Random Forest. This study used the libraries available from *scikit-learn* for Python to build the models. In terms of hyper-parameters, the default parameter values were used. These three ML algorithms were run on those seven datasets. Table 5 shows R^2 scores as the performance measures. As can be seen in Table 5, R^2 scores are very low in all models. Some are even negative. The best possible score for R^2 is 1.0 and it can be negative because the model can be arbitrarily worse (scikit-learn, 2022). As discussed earlier, the correlation coefficient between prices and review point is quite low at 0.12. The analysis shows the following for the evaluation of Scotch Whiskies from the dataset that was analyzed:

- Review points hardly explains the prices and vice versa.
- Prices and verbal descriptions are hardly related.
- Review points and verbal descriptions are minimally related.
- Verbal descriptions, categories, and prices can explain review points somewhat.

It was rather surprising to see that review points or verbal descriptions cannot explain the prices at any reasonable level. As the second and fourth rows of the Table 5 shows, R^2 values between price and review points/verbal descriptions are all negative. This implies that prices have no meaningful relationship with review points or verbal descriptions at all. As can be seen in the third row of the Table 5, review points and verbal descriptions are minimally related (R^2 from the random forest model is .2582). These facts imply that the review scores or verbal descriptions may not be beneficial to consumers for their purchasing decision making. This is quite different from wine reviews (Kim, 2022).

Conclusions

This paper analyzed the textual descriptions of whiskies as well as numeric ratings. Unlike wine sommeliers, however, the qualifications to become a whiskey sommelier are not as rigorous (Peters, 2019). As the dataset we analyzed was about the scotch whiskies, we cannot generalize our findings to other types of whiskies such as American, Irish, or others. If any datasets about those whiskies are available, it would be interesting to analyze them. Given the dataset and algorithms that were used, it was found that ratings, prices, and descriptions for Scotch whiskies are minimally consistent with each other. For purchasing decisions, the results indicate that those ratings and descriptions as well as the price are barely helpful for whiskey consumers due to their inconsistency.

References

Backhaus, A., Ashok, P. C., Praveen, B. B., Dholakia, K., & Seiffert, U. (2012). Classifying Scotch Whisky from near-infrared Raman spectra with a Radial Basis Function Network with Relevance Learning. In *ESANN*.

Fuentes, S. (2022). Implementation of Artificial Intelligence in Food Science, Food Quality, and Consumer Preference Assessment. *Foods*, *11*(9), 1192.

Gutiérrez, E. D., Dhurandhar, A., Keller, A., Meyer, P., & Cecchi, G. A. (2018). Predicting natural language descriptions of mono-molecular odorants. *Nature communications*, *9*(1), 1-12.

Kaggle, 2,2k+ Scotch Whisky Reviews. Retrieved from https://www.kaggle.com/datasets/koki25ando/22000-scotch-whisky-reviews on 4/22/2022.

Kim, B. (2022). Predictions of Wine Ratings using Natural Language Processing. *Issues in Information Systems,* 23(3), 64-68

Macias, G., Sperling, J. R., Peveler, W. J., Burley, G. A., Neale, S. L., & Clark, A. W. (2019). Whisky tasting using a bimetallic nanoplasmonic tongue. *Nanoscale*, *11*(32), 15216-15223.

Mone, G. (2020). Machine learning, meet whiskey. *Communications of the ACM*, *63*(4), 16-17.

Parker, R, Retrieved from https://en.wikipedia.org/wiki/Robert_Parker_(wine_critic)#cite_note-wa-rating-24 on 5/21/2022.

Peters, J. (2019) What is a Whiskey Sommelier? Retrieved from https://thewhiskeyjug.com/whiskey/what-is-a-whiskey-sommelier/ on 5/21/2022.

scikit-learn, Retrieved from https://scikit-learn.org/stable/modules/model_evaluation.html#r2-score-the-coefficient-of-determination on 5/21/2022.

Whisky Advocate, Retrieved from https://www.whiskyadvocate.com/ratings-and-reviews/ on 5/21/2022.