Wine Recommendation System

INSY 669 Text Analytics

Final Project

Master of Management Analytics

Desautels Faculty of Management

by

Oyundari Batbayar, Jaya Chaturvedi, Vinay Govias, Reo Paul Jackson, Moiz Shaikh, Syed Ammad Sohail

McGill University

February 13th, 2024

1   Problem Description

In an era of unprecedented choice, consumers are often overwhelmed, especially when selecting wine—a product category with over 1,500 options on average in American supermarkets. The key to navigating this vast array is personalization. A staggering 91% of consumers are drawn to brands that offer personalized recommendations (Accenture).

Our wine recommendation system cuts through the complexity, offering a user-friendly digital service that adapts to individual tastes and budgets. As e-commerce wine sales show promising growth, our system is not just a convenience but a necessity, aligning with the consumer's desire for a curated and efficient shopping experience. This innovative tool meets the modern demand for tailored experiences in the digital marketplace, positioning businesses at the forefront of the industry's evolution and driving customer satisfaction and loyalty.

2   Analytics Approach

2.1   Step 1: Text Preprocessing

Preprocessing steps for wine descriptions in our dataset include converting it to lowercase, removing punctuations and stopwords, performing lemmatization, and vectorizing the data using TF-IDF and word2vec techniques to prepare it for modeling.

2.2   Step 2: Model Development

We trained 2 types of models for our recommendation system:

- Wine Category Prediction: This text classification model predicts the wine variety (e.g., Red Blend, Riesling, Chardonnay) based on the wine's description. We tested three different algorithms: MultinomialNB, SGDClassifier, and LogisticRegression. The models were evaluated based on the hamming score and hamming loss, with LogisticRegression selected as the best model.

- Wine Label Recommendation: To recommend wine labels, we implemented a document similarity model that analyzes user input about their wine preferences (flavors, notes, aromas, etc.) and identifies the best match within our dataset. We experimented with five models - nearest neighbors using manhattan distance, euclidean distance, cosine similarity with TF-IDF vectorization and cosine similarity with

word2vec. The cosine similarity approach with TF-IDF was selected since it gave the highest accuracy for recommendations with a test dataset.

2.3   Step 3: Frontend Integration and Deployment

After experimenting and selecting our final models, we created a pipeline and saved the models using pickle files. The final recommendation system was integrated into a user-friendly frontend web application developed with Streamlit. This interface accepts user preferences regarding wines, and runs the models to generate personalized wine recommendations to users.

3   Expected Impact

The deployment of our wine recommendation system is poised to make a significant impact in three key areas.

Firstly, from a customer experience standpoint, the system will address the common issue of choice overload by delivering personalized wine selections. This targeted approach simplifies the decision-making process, allowing consumers to navigate the vast world of wine with ease and confidence.

Secondly, the system is designed to optimize sales by providing recommendations that resonate with individual customer preferences, thereby driving higher conversion rates. Such personalization fosters a more engaging shopping experience that encourages purchase completion.

Lastly, the system will serve as a rich repository of consumer preference data, offering invaluable insights that can be leveraged to inform inventory management and strategic marketing initiatives. By aligning stock with consumer tastes and market demand, retailers can operate more efficiently and market their products more effectively. In essence, this recommendation system is not just a tool for simplification but a strategic asset that contributes to a more intuitive, data-driven approach to the wine retail experience.