

Study of the relationship between competitiveness and digital footprint indicators in Valencian wineries

Leonardo Castro¹, Ana Debón¹, Josep Domenech²

¹Centro de Gestión de la Calidad y del Cambio, Universitat Politècnica de València, Spain,

²Department of Economics, Universitat Politècnica de València, Spain.

Abstract

The digital footprint of the Spanish wine sector is a valuable resource for predicting real-time indicators, enabling companies to anticipate their competitors and devise effective digital transformation strategies using emerging technologies. With advances in computation and web-scraping techniques, it is now possible to approximate competitiveness indicators using real-time information from company websites. Given this context, the general objective of this work is to analyze the relationship between the digital footprint and competitiveness of Valencian wine companies. To this end, it is proposed to use financial variables obtained from the Sistema de Análisis de Balances Ibéricos (SABI) and indicators extracted from the companies' websites. Unsupervised learning techniques will be implemented to find groups or clusters of companies based on their economic performance. Subsequently, digital footprint indicators will be used to create a supervised learning model to predict the above classification of companies based solely on digital footprint indicators to identify the most significant indicators for predicting competitiveness.

Keywords: *Digital footprint; web scraping; competitiveness; supervised and unsupervised learning; wineries.*

1. Introduction

The Spanish wine sector holds great value and relevance in the country's economy, society, and culture. According to data from the Spanish Wine Federation, Spain has approximately 13% of the total number of vineyards in the world, making it the third-largest producer and exporter of wine in terms of volume. The sector supports around 427,700 direct and indirect jobs and represents 2.2% of the Gross Value Added in Spain. Additionally, the sector has been characterized by heavy investment in innovation and development, with between 170 and 180 million euros per year invested in R&D activities over the last five years.

Due to intense competition in the market and the trend of companies toward digital transformation, the wine sector has realized the need to implement digital strategies for communication, customer acquisition, commercialization, and marketing of its products. Since digital transformation is critical to increasing business competitiveness, wine companies and wineries have increased their online presence through online sales platforms, corporate websites, and social networks, generating a significant digital footprint. This digital footprint can be detected and measured in various ways to monitor economic characteristics with real-time indicators. These indicators are valuable to private and public organizations as they enable them to anticipate competitors and identify strategies to implement within the digital transformation framework and emerging technologies (Blazquez et al. 2018).

In this context, this research aims to analyze the relationship between the digital footprint and competitiveness in Valencian wineries using multivariate techniques.

2. Materials and Methods

2.1. Data

The data matrix comprises 115 observations and 56 variables from wine companies in the Valencian Community. All companies are identified by name and their official website.

The variables used in the analysis are financial and digital footprint indicators. The financial variables were obtained from the SABI (Iberian Balance Sheet Analysis System) database, using the same variables as Rodriguez (2022). On the other hand, the digital footprint indicators used were those suggested by Blazquez and Domenech (2018), plus some others regarding company online activities on social networks.

2.2. Multivariate Analysis

Statistical analyses were conducted using the R environment for statistical computing (R Core Team, 2023). Clustering was used to identify groups of wine companies based on their financial variables, differentiating the sample of companies according to their

competitiveness characteristics. Subsequently, footprint indicators were used to create a Generalized Linear Model (GLM, McCullagh, 2019) to predict the above classification of companies using only digital footprint indicators.

Then, Receiver Operating Characteristics (ROC) graphs to evaluate the concordance between the models and actual data. Calculating the area under the ROC curve of the classifier (AUC) is a standard method to reduce ROC performance to a single scalar value representing expected performance. Therefore, as the area under the ROC curve (AUC) increased, the classifier power also increased

3. Results

After pre-processing our data, we tested clustering algorithms with different numbers of clusters (k) using the silhouette coefficient as a guide. Therefore, we decided to choose k=2 for our final analysis. We applied both K-means and fuzzy clustering algorithms to our dataset. Figure 1 shows the results obtained by applying fuzzy clustering, which resulted in two clusters with distinct financial characteristics.

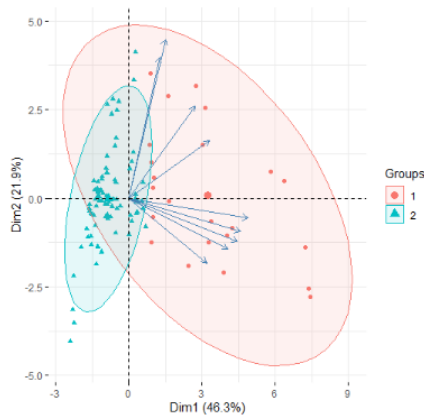


Figure 1. Results of fuzzy clustering. Source: own elaboration.

Twenty-three companies were assigned to Cluster 1, whereas 88 were assigned to Cluster 2. Cluster 1 comprises companies with higher mean values for all their competitiveness indicators; these metrics are associated with competitive performance, indicating that companies in this cluster have superior competitive performance compared to those in Cluster 2.

The logistic regression model used variables such as the age of the domain, the availability of an English version of the website, the number of Instagram posts, the number of Instagram followers, and the presence of specific keywords (such as export, efficiency,

performance, novelty, competitiveness, differentiation, LinkedIn, immaterial, brands, value, network, positioning, country, and change), using broad match based on stemming.

The model's estimated results showed a high predictive capacity, with an accuracy of 0.81 and an AUC of 0.80.

4. Conclusions

The main objective of this study was to explore the association between digital footprint indicators and competitiveness among Valencian wineries. Through exploration of the sample of companies using financial variables obtained from SABI, two clusters of companies with distinct competitiveness characteristics were identified. The first cluster consisted of companies exhibiting high levels of competitive performance and innovation capabilities, with more employees, capital, trademarks, and economic performance. On the other hand, the second cluster was composed of smaller companies with fewer employees and lower economic returns compared to the first cluster, indicating a more local market orientation.

Then, a logistic regression model was constructed, including the following variables: domain age, the English version of the website, the number of Instagram posts, the number of Instagram followers, and the presence of specific keywords. These keywords included export, efficiency, performance, novelty, competitiveness, differentiation, LinkedIn, immaterial, brands, value, network, positioning, country, and change. The model demonstrated high predictive capacity, with an accuracy of 0.81 and an AUC of 0.80.

Finally, it can be concluded that a significant relationship exists between the digital footprint indicators and a company's competitiveness, which can be used to differentiate between the previously identified clusters.

References

- Blazquez, D., Domenech, J. (2018). Web data mining for monitoring business export orientation. *Technological and Economic Development of Economy*, 24(2), 406-428.
- McCullagh, P. (2019). *Generalized linear models*. 2nd Edition. Routledge.
- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rodríguez Zamora, EE. (2022). Influencia de la actividad digital en la competitividad de las bodegas valencianas. Universitat Politècnica de València. <http://hdl.handle.net/10251/186563>.