

Prédiction de la qualité des vins du Mâconnais

Denis Collange*, Arnoult Vial*, Yoan Chabot*, Marc-Antoine Mouilleron*

*Orange Labs, France
denis.collange@orange.com

1 Introduction

Le domaine de la vitiviniculture doit répondre à de nombreux challenges ces dernières années. La météorologie (Ashenfelter et al. (2009)) et les dérèglements climatiques influencent la phénologie de la vigne et les risques de maladies, impactant le volume et la qualité des millésimes (Grifoni et al. (2006)). Par ailleurs, la concurrence et l'évolution des marchés exercent une pression non négligeable sur les acteurs du domaine. Une meilleure connaissance des consommateurs et de l'influence de la météorologie est donc vitale pour l'avenir de la filière. Dans cette optique, le projet WineCloud vise à associer les savoir-faire des vitiviniculteurs avec des techniques de gestion de données et d'IA pour aider à améliorer la qualité et la compétitivité des vins produits. Plus particulièrement, cet article expose les résultats de modèles de prédiction de la qualité des vins du Mâconnais. Plusieurs études ont cherché à comprendre les facteurs prépondérants dans la qualité des grappes, du processus de vinification et de la qualité finale du vin (Wilson (1998)).

2 Données utilisées

Plusieurs sources de données de la plateforme WineCloud ont été utilisées pour cette étude. Les données météorologiques utilisées sont celles de la station Météo France de Mâcon-Charnay, la plus proche du vignoble étudié, stockées par Infoclimat. Après nettoyage, 6 variables ont été conservées pour chaque mois : nombre de jours de précipitations et hauteur cumulée sur le mois, moyenne et minimum sur le mois des températures minimales quotidiennes, moyenne et maximum des températures maximales.

Neuf variables issues de Vivino¹ ont été extraites, pour les vins du Mâconnais : l'appellation, l'identifiant du vin, le millésime, le nombre de notes et la moyenne des notes pour ce millésime, le degré d'alcool, l'intensité gustative et la rondeur, et l'assemblage. Des grilles de dégustations détaillant les caractéristiques organoleptiques des vins (descripteurs d'arômes, propriétés olfactives et gustatives, etc.) et une estimation des durées de garde ont été fournies par un œnologue pour les 14 derniers millésimes de l'Appellation d'Origine Protégée (AOP) "Les Charmes" des Caves de Lugny. Nous avons utilisé pour ces mêmes millésimes des mesures physico-chimiques de fin de vinification : taux d'alcool volumique, de glucose-fructose, pH, acidités totale, volatile, malique et lactique, taux de gaz carbonique, SO₂ libre et SO₂ total.

1. <https://www.vivino.com/>

3 Analyse prédictive et régression linéaire

Les travaux présentés visent à prédire la qualité des vins du Mâconnais. Afin de nous acculturer à la viticulture et d'établir un dialogue avec les experts du métier, plutôt que d'essayer de trouver le modèle le plus précis, nous avons préféré commencer par utiliser un modèle de régression linéaire, dont les résultats sont facilement interprétables, pour les confronter aux connaissances empiriques issues d'une longue expérience. Nous avons utilisé ce type de modèle pour une AOP particulière, pour expliquer en fonction des données disponibles : la note moyenne attribuée par les consommateurs aux millésimes sur Vivino, la durée de garde et diverses caractéristiques organoleptiques estimées par un œnologue. Ce type d'analyse peut être utile pour adresser de manière différenciée les marchés, en fonction des préférences des consommateurs de chaque pays.

Le modèle de régression a ensuite été comparé avec d'autres algorithmes pour prédire la note moyenne des consommateurs pour un vin du Mâconnais. Nous avons observé que l'écart du modèle de régression linéaire avec les meilleures prédictions est inférieur à 1%. Nous avons notamment caractérisé l'influence de l'acidité et des températures d'été sur la qualité perçue, ou l'influence négative des températures de septembre sur la durée de garde. De nouveaux résultats ont également été découverts, par exemple que les critères gustatifs déterminants pour les consommateurs sont l'acidité et l'intensité olfactive, plutôt que les autres caractéristiques organoleptiques. Nous avons aussi pu identifier les critères sur lesquels semblent se baser les œnologues pour estimer la durée de garde. Certains de ces résultats étaient déjà connus de manière plus ou moins empirique, et avaient déjà été observés lors d'études précédentes dans des régions différentes, mais nous avons pu grâce à nos études quantifier de manière relativement précise ces corrélations, au moins dans le cas spécifique des vins du Mâconnais, avec une gamme de cépages assez réduite.

Références

- Ashenfelter, O., D. Ashmore, et R. Lalonde (2009). Bordeaux Wine Vintage Quality and the Weather. In *Collectible Investments for the High Net Worth Investor*, Volume 8, pp. 233–244.
- Grifoni, D., M. Mancini, G. Maracchi, S. Orlandini, et G. Zipoli (2006). Analysis of Italian wine quality using freely available meteorological information. *American Journal of Enology and Viticulture* 57(3), 339–346.
- Jones, G. V. et R. E. Davis (2000). Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. *American Journal of Enology and Viticulture* 51(3), 249–261.
- Kenneth Hansen (2017). Web Scraping and Analysis of Wines on Vivino.com | NYC Data Science Academy Blog.
- Van Leeuwen, C., P. Friant, X. Choné, O. Tregoat, S. Koundouras, et D. Dubourdieu (2004). Influence of climate, soil, and cultivar on terroir. *American Journal of Enology and Viticulture* 55(3), 207–217.
- Wilson, J. E. G. (1998). *Terroir*. University of California Press.