

MODELADO DEL GRADO ALCOHÓLICO COMO PARÁMETRO DE CALIDAD EN DESTILADOS DE AGAVE MEDIANTE FTIR-ATR Y APRENDIZAJE AUTOMÁTICO



Caso de estudio: destilado de agave Flammam

Maestría en Ciencia de Datos · Facultad de Ciencias Físico Matemáticas · Universidad Autónoma de Nuevo León

GARCÍA-PÉREZ Luis Daniel **BENAVIDES-VÁZQUEZ José Alberto** **NOCEDO-MENA Deyani** **RÍOS-MERCADO Azucena Yoloxóchitl**

lgarciape@uanl.edu.mx jbenavidesv@uanl.edu.mx deyani.nocedomn@uanl.edu.mx azucena.riosmr@uanl.edu.mx



INTRODUCCIÓN Y PLANTEAMIENTO DEL PROBLEMA

El control de calidad de los destilados de agave depende de métodos de referencia (cromatografía de gases, volumetría) que consumen horas, reactivos y muestra: inviable para el monitoreo rutinario de un productor artesanal como Flammam.

La espectroscopía FTIR-ATR mide en minutos, sin preparación de muestra y sin destruirla. El reto no es medir: es traducir esa huella espectral en un valor de calidad confiable. El grado alcohólico es el parámetro regulado más relevante (35 a 55 % v/v; SE, 2012; SE, 2016) y el etanol absorbe con bandas definidas en el infrarrojo medio (Rohman y Che Man, 2012).

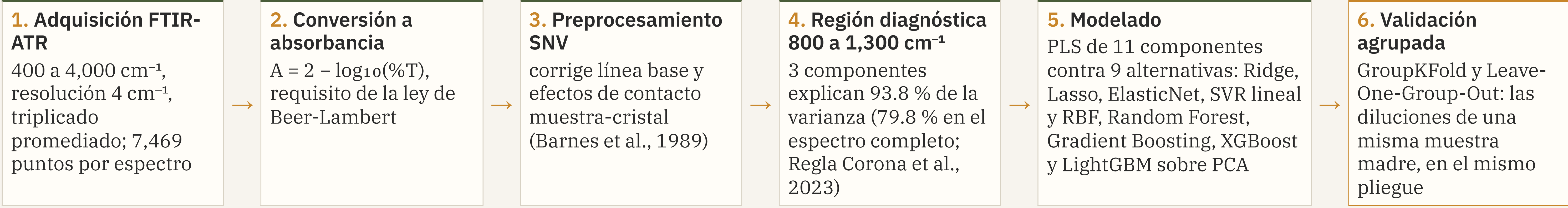
OBJETIVO Y PREGUNTA

Objetivo: Desarrollar y evaluar modelos de aprendizaje automático basados en espectros FTIR-ATR para estimar el grado alcohólico en destilados de agave. Caso de estudio: destilado Flammam.

Pregunta de investigación: ¿Puede el espectro FTIR-ATR predecir el grado alcohólico con error bajo para monitoreo preliminar, bajo validación libre de fuga de información?

METODOLOGÍA

Conjunto de datos: 46 muestras —22 puras y 24 diluciones derivadas de 5 muestras madre— en un rango de 22 a 56 % v/v. Python 3.13 con numpy, pandas, matplotlib y scikit-learn (Pedregosa et al., 2011).



RESULTADOS

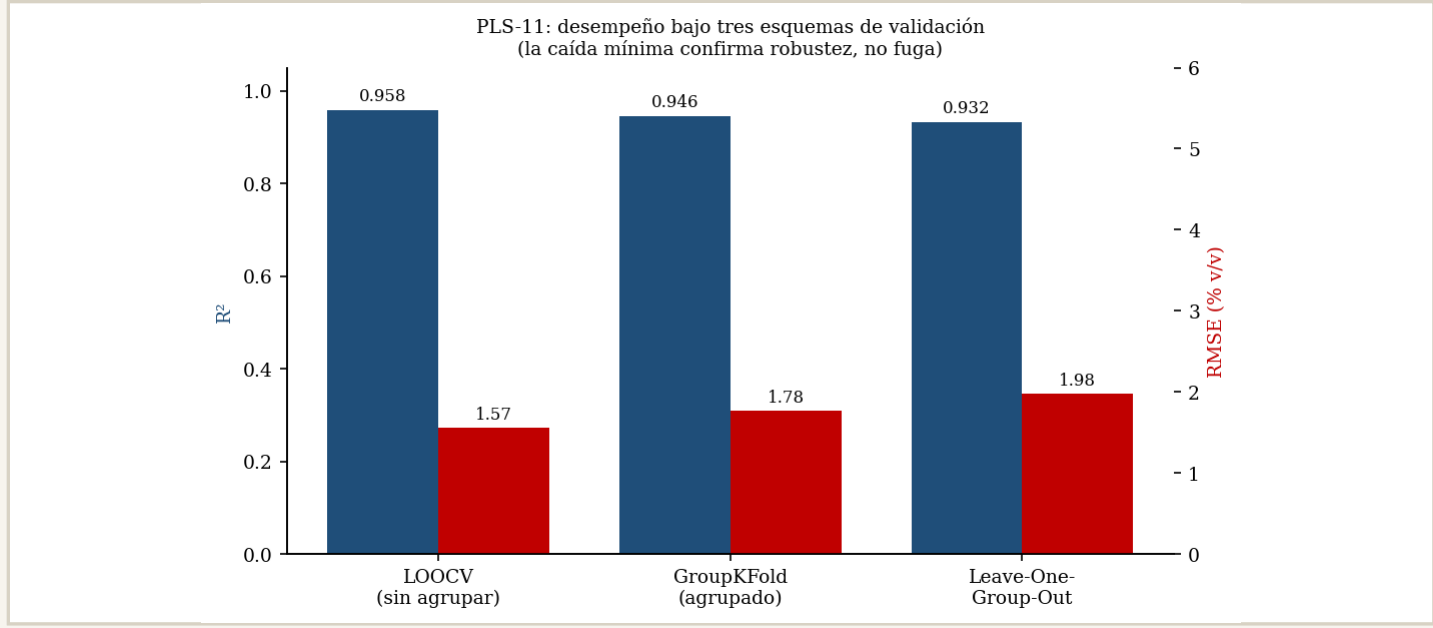


Figura 1. PLS bajo tres esquemas de validación de exigencia creciente.

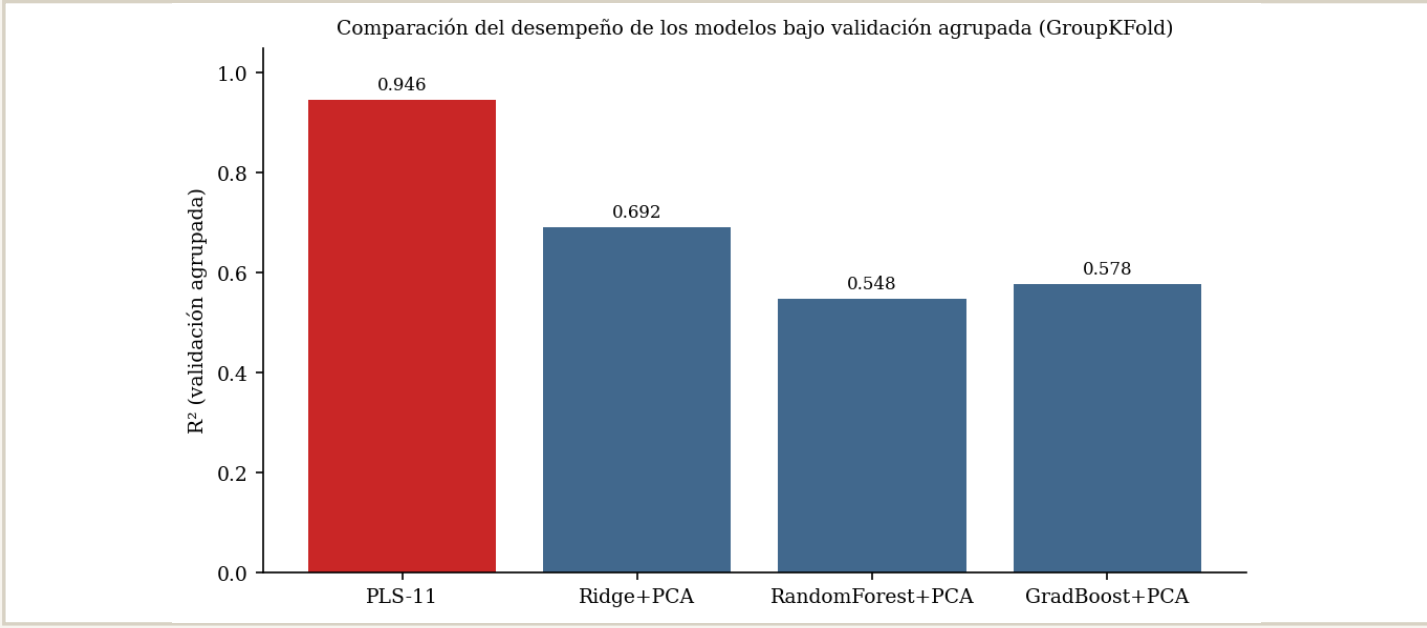


Figura 2. Los 10 modelos bajo idéntica validación agrupada (GroupKFold). El PLS supera al segundo mejor por 0.248 en R^2 .

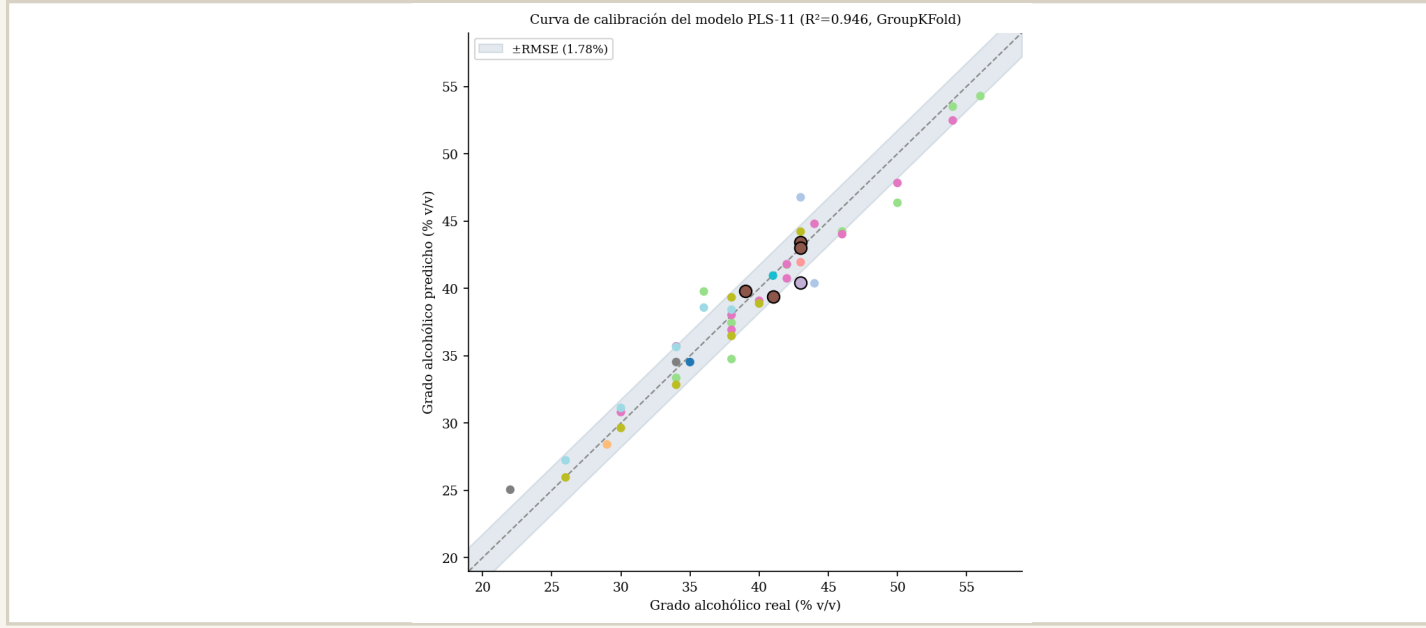


Figura 3. Curva de calibración del PLS: predicho frente a real, con banda de $\pm \text{RMSE} = 1.78$ % v/v.

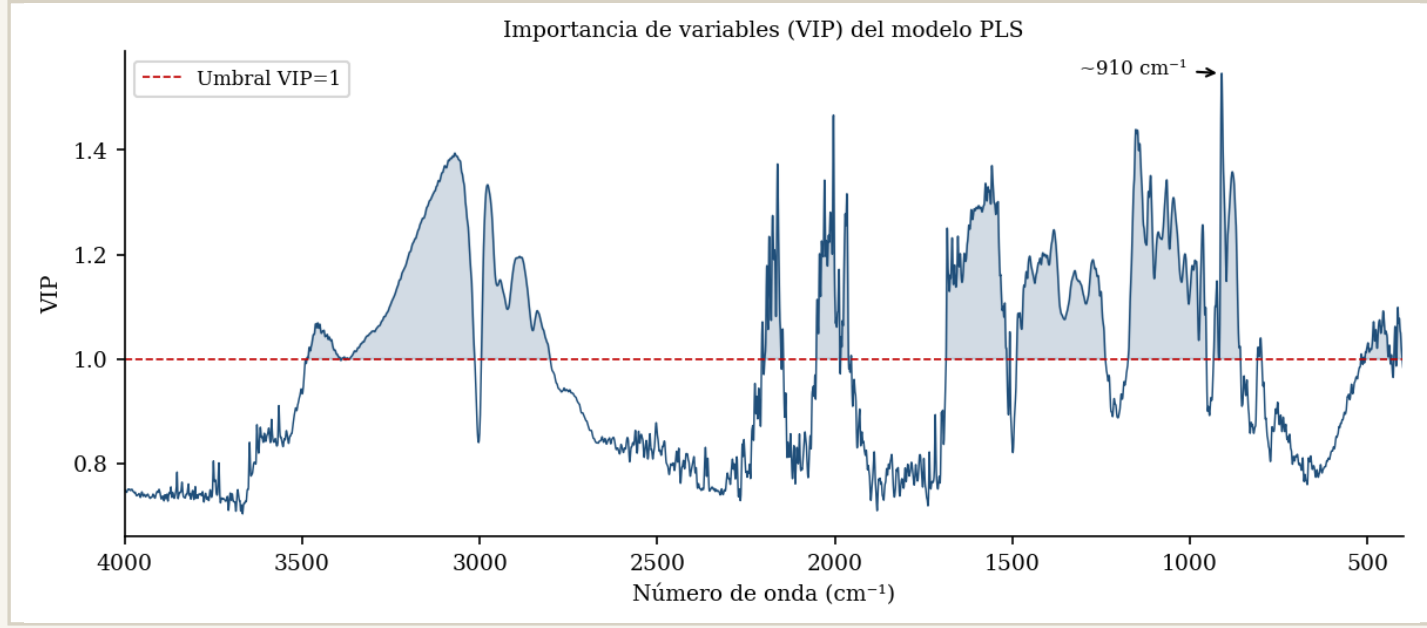


Figura 4. Índice VIP del PLS. Los máximos coinciden con la banda C–O del etanol ($\sim 1,150 \text{ cm}^{-1}$) y los carbohidratos residuales ($\sim 910 \text{ cm}^{-1}$).

Tabla 1. Desempeño bajo validación agrupada

Modelo	R^2	RMSE	MAE
PLS (11 comp.)	0.946	1.78	1.40
SVR-RBF + PCA	0.698	4.19	2.55
Ridge + PCA	0.692	4.23	2.42
LightGBM + PCA	−0.050	7.81	6.02

Nota: RMSE y MAE en % v/v. Se omiten 6 modelos intermedios.

Tabla 2. Desempeño del PLS por esquema de validación

Esquema	R^2	RMSE
LOOCV sin agrupar	0.958	1.57
GroupKFold	0.946	1.78
Leave-One-Group-Out	0.932	1.98

Nota: RMSE en % v/v.

El PLS superó a todos los modelos de comparación bajo idéntica validación, algo coherente con la estructura del problema: muchas más variables que muestras ($p = 7,469 \gg n = 46$) y relación absorbancia-concentración lineal, por lo que los ensambles de árboles sobreajustan sin ganancia (Wold et al., 2001).

Al endurecer la validación el R^2 cae solo 0.026: el desempeño se sostiene en una calibración redundante que cubre todo el rango (Bevilacqua et al., 2012), no en fuga de información. El VIP ubica las bandas más informativas justo donde absorbe el etanol.

¿Por qué hay que agrupar las diluciones?

Sin agrupar (LOOCV)



El modelo ya vio espectros casi idénticos de la misma muestra. Mide memorización.

$R^2 = 0.958$

Agrupado (Leave-One-Group-Out)



Ninguna versión de la muestra evaluada estuvo en entrenamiento. Mide generalización.

$R^2 = 0.932$

Figura 5. La diferencia entre ambos esquemas es de 0.026 en R^2 . Esa estabilidad, y no la cifra alta, sostiene el trabajo.

CONCLUSIONES

DESEMPEÑO

$R^2 = 0.946$

El flujo FTIR-ATR $\rightarrow$ absorbancia $\rightarrow$ SNV $\rightarrow$ PLS estima el grado alcohólico por encima del umbral de aplicabilidad de la literatura ($R^2 > 0.90$; $\text{RMSE} < 2$ % v/v).

Posición frente a la literatura: los estudios previos reportan caracterización exploratoria por PCA (Regla Corona et al., 2023) o calibración PLS con validación externa simple (Lachenmeier, 2007). Este trabajo documenta la calibración bajo agrupamiento por muestra de origen.

REFERENCIAS

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777.

Regla Corona, D. N., Mondragón Cortez, P. M., & López Ramírez, J. E. (2023). Huella espectral de bebidas destiladas de agave mediante espectroscopía de infrarrojo. En A. C. Gschaedler Mathis (Ed.), *Valorización de bagazo de agave mediante la creación de biorefinerías* (pp. 199–207). CIATEJ.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Bevilacqua, M., Bro, R., Marini, F., & Rinnan, Å. (2012). Recent trends in the application of chemometrics to foodomics. *Trends in Analytical Chemistry*, 40, 3–12.

Rohman, A., & Che Man, Y. B. (2012). Fourier transform infrared (FTIR) spectroscopy for analysis of food adulteration. *Applied Spectroscopy Reviews*, 47(1), 1–20.

Secretaría de Economía. (2012). *NOM-006-SCFI-2012. Bebidas alcohólicas – Tequila – Especificaciones*. Diario Oficial de la Federación.

Lachenmeier, D. W. (2007). Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. *Food Chemistry*, 101(2), 825–832.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

Secretaría de Economía. (2016). *NOM-070-SCFI-2016. Bebidas alcohólicas – Mezcal – Especificaciones*. Diario Oficial de la Federación.