




Enseñanza del *machine learning* y la quimiometría en química analítica mediante propuestas prácticas e interactivas

Teaching of Machine Learning and Chemometrics in Analytical Chemistry Based on Interactive Hands-on Activities

Ángel Sánchez-Illana^a, Bayden R. Wood^b y David Pérez-Guaita^{c*}

^aDepartamento de Química Analítica, Facultad de Química, Universitat de València 

^bCentre for Biospectroscopy and School of Chemistry, Monash University 

^cDepartamento de Química Analítica, Facultad de Química, Universitat de València, david.perez-guaita@uv.es 

How to cite: Ángel Sánchez-Illana, Bayden R. Wood y David Pérez-Guaita. 2023. Enseñanza del machine learning en química analítica mediante propuestas prácticas e interactivas. En libro de actas: *IX Congreso de Innovación Educativa y Docencia en Red*. Valencia, 13 - 14 de julio de 2023. Doi: <https://doi.org/10.4995/INRED2023.2023.16679>

Abstract

Despite the growing popularity of machine learning (ML), the teaching of this disruptive field in analytical chemistry is challenging due to the lack of enough programming background in both, professors, and students. Because of that, this subject is sometimes underrated or even ignored in chemistry curriculums. In this work, we firstly surveyed the previous knowledge in multivariate analysis and programming by students enrolled in the master's degree in chemistry. Upon recognizing a deficiency in fundamental programming and statistical principles, we carried out actions to close the gap between ML and analytical chemistry in under- and post-graduate level. Accordingly, we proposed the use of the interactive software Orange and the programming of apps with MATLAB for teaching ML in the laboratory lessons of analytical chemistry. With this approach, two laboratory lessons were designed and conducted which are focused on analysis of foodstuffs by infrared spectroscopy and using ML in daily contexts. The evaluation of the methodologies proposed indicated that the use of interactive software made ML more appealing to the students and contributed to a better understanding of ML concepts.

Keywords: Machine Learning, Analytical Chemistry, Chemometrics, Methodology, Laboratory, Programming

Resumen

A pesar de la creciente popularidad de las técnicas de aprendizaje automático (“machine learning”, en inglés) su enseñanza en química analítica es un reto debido a la falta de conocimientos en programación por parte del alumnado y el profesorado. Debido a ello, estas técnicas suelen obviarse o incluirse sucintamente en los currículos. En este trabajo, estudiamos los conocimientos previos de programación y análisis multivariante del alumnado matriculado en el Máster en Química e identificamos diferentes deficiencias en conceptos básicos de programación y estadística. En consecuencia, para cerrar la brecha entre el aprendizaje automático y la química analítica, planteamos el uso del programa

interactivo Orange y la programación de aplicaciones con MATLAB. Utilizando este enfoque, diseñamos y llevamos a cabo dos sesiones prácticas consistentes en el análisis de alimentos mediante espectroscopía infrarroja y en la implementación de modelos de aprendizaje automático aplicados a contextos cotidianos. Tras evaluar las metodologías propuestas, comprobamos que estas hacen el aprendizaje automático más atractivo para el estudiantado contribuyendo a su mejor aprendizaje.

Palabras clave: *Aprendizaje Automático, Machine Learning, Química Analítica, Quimiometría, Metodología, Laboratorio*

1. Introducción

En las últimas décadas, los avances en computación y análisis de datos han transformado profundamente muchos aspectos de la química. En este contexto, la quimiometría y el aprendizaje automático (más conocido por su denominación en inglés, *machine learning*) juegan un papel clave en el análisis masivo de datos adquiridos mediante los instrumentos y sensores de última generación (Debus et al., 2021; Joshi, 2023). Estos avances han conducido a una creciente preocupación acerca de la obsolescencia del currículo universitario de química analítica. En particular, se ha sugerido que los programas universitarios necesitan adaptarse para asegurar que los graduados estén preparados para trabajar en una sociedad cada vez más digitalizada y automatizada, en la que la capacidad de aprovechar las herramientas de computación e inteligencia artificial es cada vez más importante (Holme, 2019).

Por otra parte, desde hace ya más de medio siglo se lleva observando la dificultad para introducir la quimiometría y el *machine learning* en el currículo de química analítica (Howery & Hirsch, 1983). En los primeros estudios que se realizaron al respecto, se identificó que una de las principales limitaciones era la carencia de conocimientos previos de programación informática. Para abordar este problema se diseñaron cursos que incluían una formación básica en programación y se empleaban ejemplos prácticos utilizando computadores de la época. Uno de los primeros cursos de este tipo reportados en la literatura fue el realizado en la *Tufts University* de Medford (EE. UU.) en el que el alumnado empleaba un enorme computador DECsystem-10 (Figura 1) y programación en FORTRAN para implementar diferentes algoritmos trabajados previamente en clases teóricas (Delaney & Warren, 1981).



Figura 1. Computador DECsystem-10. Fuente: *Retro-Computing Society of Rhode Island*

Afortunadamente, hoy en día existen muchísimas más facilidades para poder desarrollar este tipo de cursos ya que la potencia de cálculo de cualquier ordenador personal actual es suficiente para llevar a cabo algoritmos de aprendizaje automático y quimiométricos con grandes conjuntos de datos. Además, existen multitud de programas propietarios y de código abierto disponibles, destacando aquellos escritos en lenguajes interpretados de alto nivel como Python, R, o MATLAB. Cada lenguaje ofrece sus pros y contras

pero comparten grandes ventajas en cuanto a su uso para aprender a programar en contextos en los que se realice intensivamente la manipulación de vectores y matrices, el procesado de señales, el uso de estadística avanzada y la visualización de datos (Fangohr, 2004; Tadayon, 2020). Esto hace que sean lenguajes ideales para la enseñanza del *machine learning* y la quimiometría lo que ha dado lugar a diferentes propuestas de innovación educativa que han generado creciente evidencia al respecto (St James et al., 2023).

Empleando estos lenguajes pueden programarse aplicaciones interactivas con interfaces gráficas como el software comercial PLS_Toolbox (*Eigenvector Research .Inc*, Manson, WA, EE. UU.) programado en MATLAB y el programa en código abierto Orange, escrito en Python (Demsar et al., 2013). También cabe destacar el complemento *App Designer* de MATLAB con el que es posible crear aplicaciones interactivas con relativa facilidad. Como puede verse en la Figura 2, tanto PLS_Toolbox, Orange y aplicaciones creadas con MATLAB *App Designer* son muy intuitivas frente a un script de MATLAB o Python.

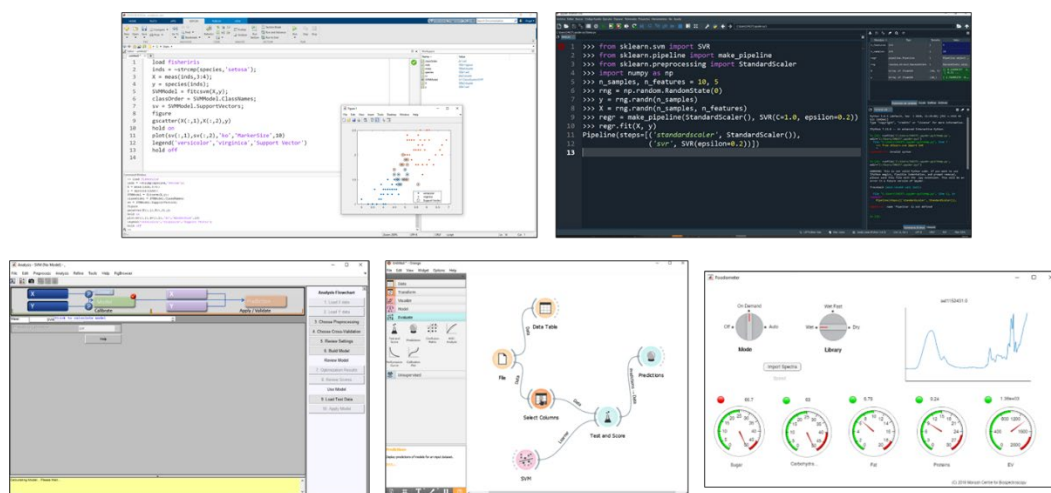


Figura 2. Script de MATLAB (arriba izquierda) y Python (arriba derecha) con un algoritmo de aprendizaje automático; interfaz del programa PLS_Toolbox (abajo izquierda); interfaz del programa Orange (abajo centro); interfaz de un programa creado con MATLAB App Designer (abajo derecha)

En este trabajo, trasladamos estas innovaciones a los estudios en química analítica de la Universidad de Valencia incluyendo en unas sesiones prácticas el uso del programa Orange y la programación *ad hoc* de aplicaciones en MATLAB. En estas prácticas analizamos datos obtenidos en el contexto del análisis de alimentos y pusimos a punto métodos de *machine learning* empleados en recientes publicaciones científicas (Perez-Guaita et al., 2021). Asimismo, evaluamos los resultados obtenidos en esta innovación realizando encuestas antes y después de estas prácticas.

2. Objetivos

Los objetivos principales de esta propuesta son:

- Identificar las carencias en programación, *machine learning* y quimiometría de los alumnos matriculados en el Máster en Química de la Universidad de Valencia.
- Enseñar la gran utilidad del *machine learning* para el análisis de datos químicos y hacer más atractiva la materia para el alumnado.
- Comparar dos programas para la enseñanza del *machine learning* en química. Por un lado, Orange, un software intuitivo y por otro MATLAB, un lenguaje de programación. En ambos casos se enseñará su uso a través de ejemplos analíticos con un set de datos de espectros de IR de alimentos.

3. Desarrollo de la innovación

Las sesiones prácticas para llevar a cabo esta innovación se realizaron durante el curso 2021/2022 en la asignatura de “Estrategias analíticas para la resolución de problemas socioeconómicos” que forma parte del Máster Universitario en Química de la Universitat de València. Esta asignatura tiene entre sus competencias la enseñanza del tratamiento quimiométrico de datos como herramienta para la obtención de información de calidad. Por lo tanto, hipotetizamos que esta innovación podría tener un impacto sustancial en la adquisición de competencias y en los resultados de aprendizaje relacionados con este punto.

El desarrollo de la práctica se dividió en 2 sesiones en días diferentes incluyendo en estas sesiones el análisis exploratorio de los espectros, el análisis mediante técnicas de aprendizaje automático, la discusión de los resultados y las encuestas previas y finales. Ambas sesiones se realizaron posteriormente a la parte teórica de la asignatura en la que se habían explicado los fundamentos de la quimiometría y el aprendizaje automático en el contexto de la química analítica.

3.1. Encuesta de conocimientos previos

Como primer paso, al inicio de la asignatura (es decir, previamente a las clases teóricas correspondientes), se realizó una encuesta utilizando la aplicación Formularios de Google. Las preguntas se seleccionaron para sondear los conocimientos previos y el interés del alumnado en quimiometría y aprendizaje automático. Se realizaron preguntas de respuesta abierta, en las que podía desarrollarse cualquier explicación sin limitación, y otras del tipo acuerdo-desacuerdo en las que había que escoger entre muy en desacuerdo, en desacuerdo, neutral, acuerdo y muy de acuerdo. Las preguntas escogidas se encuentran en la Tabla 1.

Tabla 1. Preguntas del cuestionario inicial. En las preguntas de respuesta abierta se permitía responder libremente cualquier texto, en las preguntas de acuerdo-desacuerdo había que elegir entre muy en desacuerdo, en desacuerdo, neutral, de acuerdo o muy de acuerdo.

Tipo de preguntas	Pregunta
	¿Conoces algún método de análisis multivariantes o quimiométrico? ¿Cuáles?
	¿Conoces algún método de Machine Learning (ML)? ¿Cuáles?
	¿Cómo ha sido tu formación en métodos quimiométricos?
Respuesta abierta	¿Has recibido formación sobre métodos de ML? ¿Cuáles?, ¿fue intensiva o superficial?
	¿Has utilizado alguna vez algún método de análisis multivariante o quimiométrico? ¿Cuáles?
	¿Has utilizado alguna vez algún método de ML (creado un modelo)? ¿Cuál?
	¿Qué software para realizar ML o quimiometría conoces?
	¿Qué software para realizar ML o quimiometría has utilizado?
Acuerdo-desacuerdo	La quimiometría trata los datos químicos, mientras que el ML se refiere al tratamiento de datos de otras disciplinas
	Los algoritmos de ML son demasiado complejos para poder usarse con datos químicos
	La quimiometría y el ML son conceptos relacionados ya que se basan en el uso de estadística multivariante.

La quimiometría usa estadística multivariante mientras que el ML usa algoritmos informáticos avanzados.

La quimiometría la puede hacer un químico con software dedicado, para realizar el ML se necesita un alto nivel de programación

La quimiometría o el ML pueden ser útiles en el campo de química analítica

Me interesa aprender quimiometría

Me interesa aprender Machine Learning

3.2. Sesiones prácticas

Las sesiones se realizaron en un aula de informática de PC, equipada con ordenadores con el sistema operativo Windows 10 a los que previamente se les había instalado MATLAB y Orange. En todas las sesiones prácticas el profesor, primeramente, realizaba una introducción con diapositivas explicativas en el ordenador del aula para hacer un repaso de los algoritmos que se iban a utilizar (previamente explicados en las sesiones teóricas) junto con unos fundamentos de MATLAB y Orange. Posteriormente, el profesor realizaba una serie de ejercicios como ejemplo que, a continuación, se repetían para que los alumnos pudieran seguirlos paso a paso junto al profesor a tiempo real. Por último, se proponían una serie de ejercicios que los alumnos hacían independientemente con la asistencia del profesor si era necesario. Los ejercicios se realizaban primero en MATLAB y después en Orange. Los sets de datos utilizados en las sesiones prácticas fueron obtenidos de repositorios públicos, de los ejemplos proporcionados con los propios programas y de un conjunto de espectros adquiridos por alumnos de estudios de química de la *Monash University* de Melbourne (Australia). Al alumnado se le proporcionaron los programas y scripts necesarios para seguir las clases mediante el aula virtual. Todo este material está disponible previa petición razonada al autor de correspondencia.

3.2.1. Sesión 1. Análisis exploratorio (4 horas)

En una primera sesión, tras la realización de una serie de ejercicios de visualización de datos elemental con set de datos públicos, se realizaron ejercicios de visualización de espectros utilizando los datasets de alimentos de la *Monash University*. Se utilizaron estos ejemplos para repasar conceptos de espectroscopía vibracional y asignación de bandas. Después, se pasó a realizar ejercicios de análisis no supervisado mediante análisis de componentes principales (PCA, del inglés *principal component analysis*).

3.2.2. Sesión 2. Uso de metodologías de aprendizaje automático para la predicción de parámetros nutricionales (4 horas)

En una segunda sesión se implementaron diferentes modelos de análisis supervisado mediante el uso de los algoritmos de mínimos cuadrados parciales (PLS, del inglés *partial least squares*), random forest y de máquinas de vectores de soporte (SVM, del inglés *support vector machines*). Se construyen diferentes modelos para clasificación y para predicción. Estos ejercicios prácticos se contextualizaron en el marco de los conceptos explicados en las clases teóricas (recordando los conceptos de sobreajuste y diferente tipos de errores y validación, entre otros).

Finalmente, para ilustrar el uso del MATLAB *App Designer* y las aplicaciones de MATLAB en la implementación de modelos de machine learning, se les proporciona a los alumnos un código de una aplicación de MATLAB diseñada para el análisis de alimentos basada en un trabajo de investigación

publicado recientemente (Pérez-Guaita, 2021). Se les proporcionan a los estudiantes diferentes sets de datos de espectros adquiridos en diferentes condiciones de alimentos para que evalúen el desempeño de la aplicación.

3.3. Encuesta de conocimientos adquiridos

Finalmente, tras la última sesión se realiza la encuesta final con una serie de preguntas seleccionadas para valorar la utilidad del enfoque y para que el alumnado valore comparativamente MATLAB y Orange. Las preguntas concretas se encuentran detalladas en la Tabla 2.

Tabla 2. Preguntas del cuestionario inicial. En las preguntas de respuesta abierta se permitía responder libremente cualquier texto, en las preguntas de acuerdo-desacuerdo había que elegir entre muy en desacuerdo, en desacuerdo, neutral, de acuerdo o muy de acuerdo.

Tipo de preguntas	Pregunta
Acuerdo-Desacuerdo	¿Te han sido útil los scripts de Matlab a la hora de ENTENDER cómo funcionan los métodos de tratamiento de datos de la asignatura (PCA,PLS...)?
	¿Te ha sido útil Orange a la hora de ENTENDER cómo funcionan los métodos de tratamiento de datos de la asignatura (PCA, PLS...)?
	¿Te han sido útil los scripts de Matlab a la hora de APLICAR los métodos de tratamiento de datos de la asignatura (PCA,PLS...)?
	¿Te ha sido útil Orange a la hora de APLICAR los métodos de tratamiento de datos de la asignatura (PCA,PLS...)?
	¿Qué software crees que es más útil para, en un futuro, desarrollar aplicaciones reales?
Respuesta abierta	¿Serías capaz de utilizar, hoy, el software Matlab para tratar datos de tu TFM?
	¿Serías capaz de utilizar, hoy, el software Orange para tratar datos de tu TFM?
	¿Qué ventajas/desventajas le ves al Matlab como método de aprendizaje de quimiometría?
	¿Qué ventajas/desventajas le ves al orange como método de aprendizaje de quimiometría?
	¿Qué ventajas/desventajas le ves al Matlab como método de tratamiento de datos?
¿Qué ventajas/desventajas le ves al Orange como método de tratamiento de datos?	
En un futuro si tuvieras más tiempo, ¿Sobre qué software te gustaría aprender más?	
Comenta, si quieres, cualquier feedback que crees que podría mejorar la asignatura.	

Resultados

4.1. Encuesta de conocimientos previos

La Figura 3 muestra los resultados de las preguntas de la encuesta inicial. Se puede observar que los estudiantes reportan mínimos conocimientos cuando se les pregunta sobre “Métodos quimiométricos” o “Análisis multivariante”. Si bien es cierto que conocen el PCA, más del 90% afirman que su formación ha sido superficial o no han recibido formación. Respecto a la palabra clave “*machine learning*” los alumnos

dicen reconocer ningún método. En este contexto, los alumnos también declaran que no han utilizado nunca ningún software para realizar ningún modelo ni quimiométrico ni de *machine learning*.

Respecto a los conceptos “Quimiometría” y “*machine learning*”, los estudiantes no muestran mucho conocimiento sobre lo que son y significan. Sólo un 34% de los estudiantes estuvo de acuerdo o muy de acuerdo con la afirmación “La quimiometría y el *machine learning* son conceptos relacionados ya que ambos se basan en la estadística multivariante”. Sin embargo, parece ser que los estudiantes sí que intuyen cierta aplicabilidad del *machine learning* y la quimiometría en la química: Sólo el 8% estaba de acuerdo con la afirmación “Los algoritmos de Machine Learning son demasiado complejos para la química analítica” y en desacuerdo con “La quimiometría o el ML pueden ser útiles en el campo de química analítica”. Además, más del 50% estaban en muy de acuerdo o de acuerdo con las afirmaciones “Me interesa aprender *machine learning*/Quimiometría”.

Los resultados de la encuesta claramente señalan una falta de conceptos relativos al tratamiento multivariante y su alto potencial en el campo de la química analítica. En el currículo del grado en química en la Universitat de València, de donde provenía la mayoría del alumnado, se imparte un tema de estadística multivariante, estudiando por ejemplo métodos como el PCA o el PLS en la asignatura química analítica III, pero parece ser que los alumnos no son capaces de asumir estos conocimientos. Un posible factor es la ausencia de clases prácticas, lo cual está confirmado por el hecho de que ningún alumno había realizado ningún modelo ni utilizado ningún software. Esto también se ve reflejado en la vaguedad con que los alumnos contestaron a las preguntas relativas a las diferencias entre “*Machine Learning*” y quimiometría. A pesar de esto, los resultados también indican una gran predisposición de la mayoría de los alumnos a aprender este tipo de análisis.

En estas circunstancias se plantearon ciertas dificultades a la hora de obtener las competencias previstas en la guía docente de la asignatura, donde se habla de “Seleccionar y aplicar, entre las principales técnicas quimiométricas de análisis multivariante, aquella/s que resulte/n más adecuada/s para el tratamiento de datos analíticos complejos, e interpretar adecuadamente los resultados obtenidos.” Por ello se diseñó una metodología docente con tres objetivos claros. El primero, el refuerzo de conceptos básicos de estadística multivariante, así como de los métodos supervisados (PCA, clúster análisis) y no supervisados (PLS, SVM). El segundo, que los estudiantes utilizaran de forma práctica software dedicado para poder realizar los modelos y el tercero que pudieran aplicar esos modelos con aplicaciones reales para estudiar el potencial y al valor de estas técnicas.

La enseñanza del machine learning mediante propuestas prácticas e interactivas: la programación de aplicaciones en MATLAB para el análisis de datos espectrales de alimentos

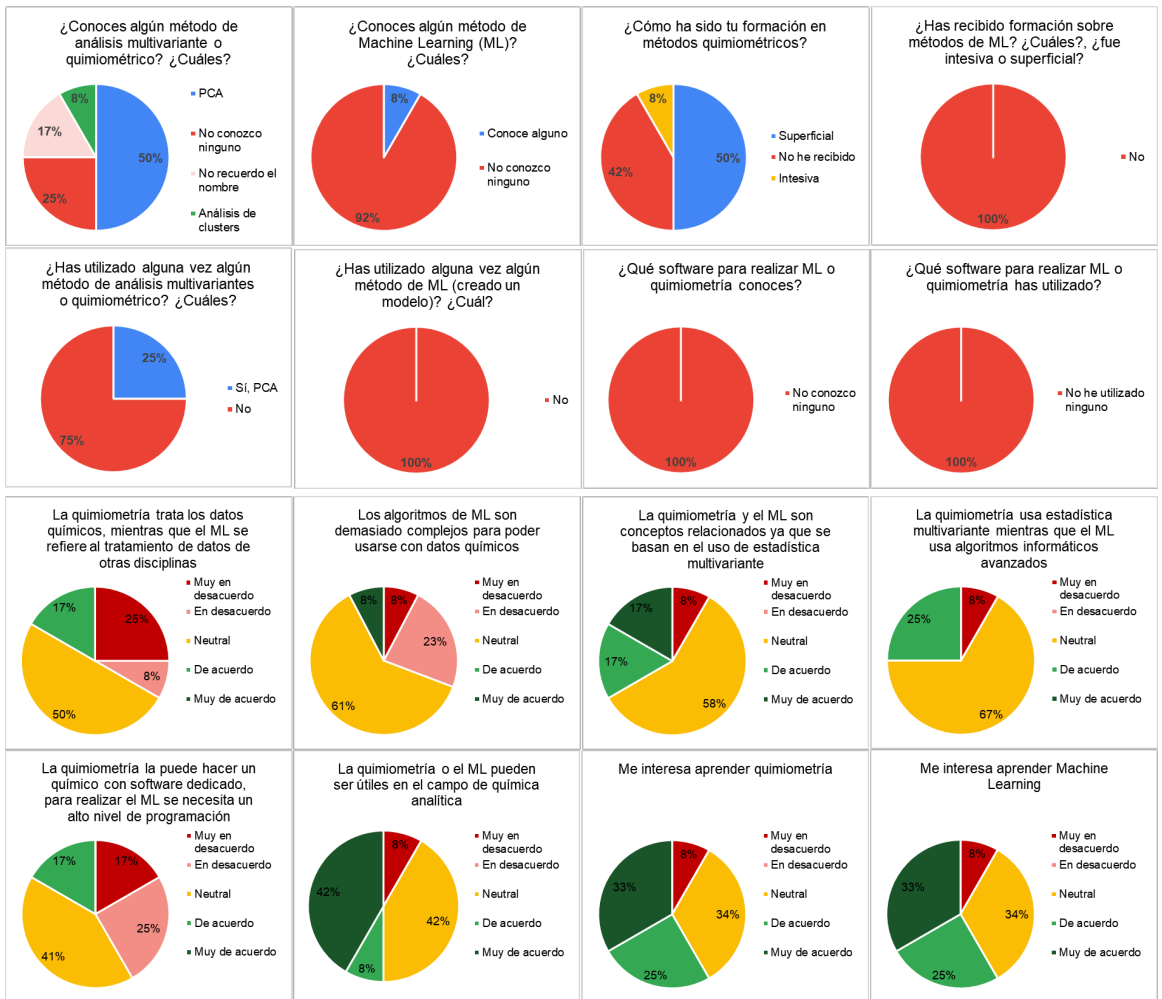


Figura 3. Resultados de la encuesta inicial.

4.2. Desarrollo de las sesiones de prácticas

4.2.1 Análisis visual

Las sesiones de prácticas se realizaban inmediatamente después de clases magistrales donde se presentaban conceptos teóricos. Primero se presentó el set los espectros de IR de alimentos. Tenía la ventaja de que los estudiantes ya conocen, por cultura general, los mayores componentes de los alimentos. Esto ayudó a que los alumnos pudieran reconocer y asignar con facilidad sus bandas de IR de. En la Figura 4a se pueden observar los diferentes espectros, destacando por ejemplo a los alcoholes de los carbohidratos en el sirope (flecha verde) a las amidas de las proteínas en las proteínas en polvo (flecha roja) o los enlaces C-H en el aceite (flecha azul).

4.2.1 Análisis no supervisado

Una vez los alumnos estaban familiarizados con los espectros, se realizó el análisis no supervisado con MATLAB y Orange. Los alumnos pueden aplicar sus conocimientos en PCA a los datos, obteniendo clúster de las muestras en los *scores* como el que se observa en la Figura 4b. Se puede observar que hay un grupo de muestras con valores positivos del componente principal 1. Estas muestras son las que contienen mucha

agua y no están secas como las leches y los zumos. Los alumnos pueden relacionar ahora ese hecho con los *loadings* del PC1, que tiene bandas positivas para las típicas bandas del agua (1600 y $3000-3500\text{ cm}^{-1}$). Además, el PC2 es capaz de diferenciar alimentos con mucho azúcar (sirope) de alimentos con muchas proteínas (proteínas en polvo). Los *loadings* del PC2 lo confirman mostrando bandas positivas para los alcoholes (1000 cm^{-1}) y negativas para las amidas $1400-1600\text{ cm}^{-1}$.

Con esto, los alumnos son capaces de relacionar, en un escenario real y conocido, la relación que existe entre las muestras, las variables, los scores y los loadings. También se familiarizan con los programas y entienden a interpretar la información que se obtiene. Se hace hincapié en la posibilidad de realizar este tipo de análisis multivariante en una infinidad de set de datos (datos económicos de países, estadísticas deportivas, etc.).

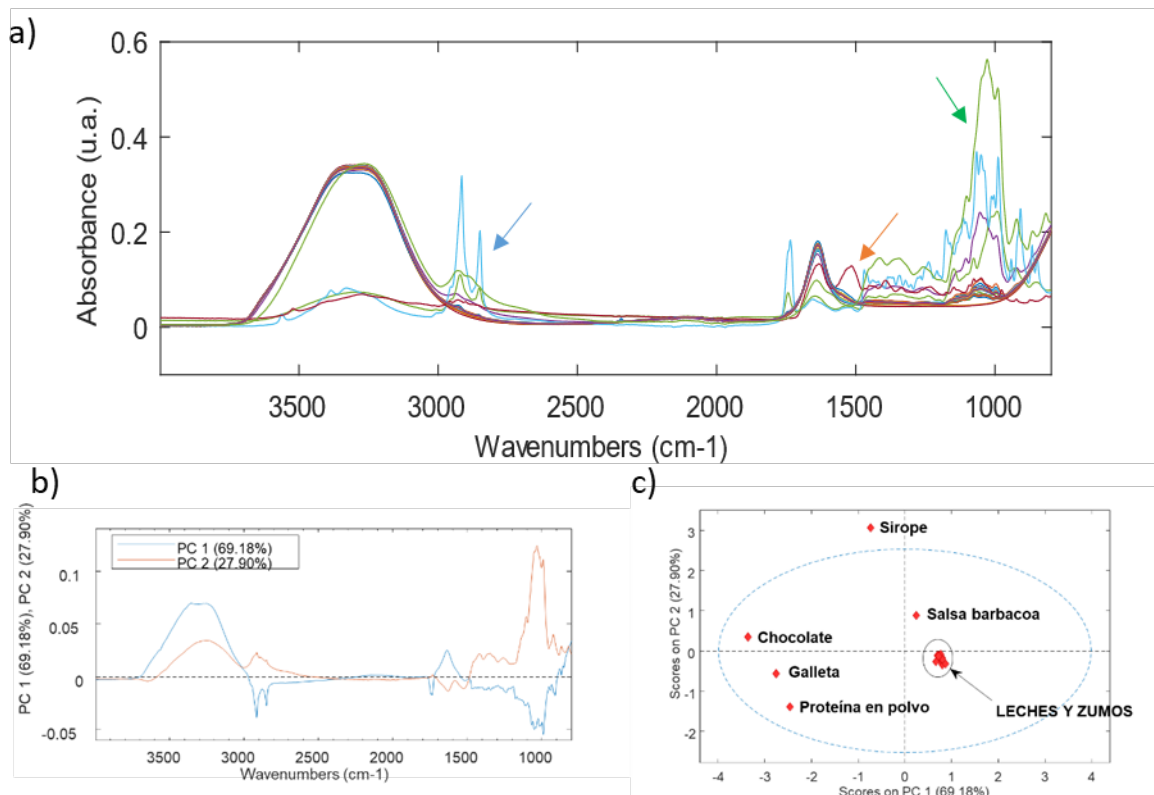


Figura 4. a) Espectros IR de diferentes alimentos obtenidos por alumnos durante la sesión de laboratorio. Resultados del PCA loadings (b) y scores (c)

4.2.1 Análisis Supervisado e implementación en una app de Matlab.

A continuación, y tras haber realizado clases magistrales donde se introduce el análisis supervisado (regresiones y clasificaciones), se realizan modelos cuantitativos (PLS, SVM, y random forest) para predecir la cantidad de los diferentes parámetros nutricionales con los espectros de IR. El alumnado trabaja la necesidad de dividir los sets en calibración y test para poder evaluar los errores de generalización de forma correcta. Además, calculan errores de calibración y validación cruzada con modelos de diferente complejidad (por ejemplo, con diferente número de variables latentes) para estudiar conceptos como el *overfitting* y el *underfitting*.

Finalmente, para mostrar a los alumnos la facilidad con la que este tipo de modelos predictivos puede ser implementada, se utiliza una app de MATLAB desarrollada previamente para la cuantificación de parámetros nutricionales en alimentos (Pérez-Guaita, 2021). Una captura de la app se puede observar en la Figura 5, donde se aprecia como, tras importar un espectro de un alimento, la app lo procesa y devuelve, en unos segundos, una serie de predicciones sobre la cantidad de diferentes parámetros nutricionales. Se explica a los alumnos que este tipo de apps requieren para su desarrollo la utilización de lenguajes de programación como Matlab, pero para su no es necesario y que además es muy intuitivo. Con ello se pone en valor el uso de estadística multivariante y se anima a utilizar lenguajes de programación.

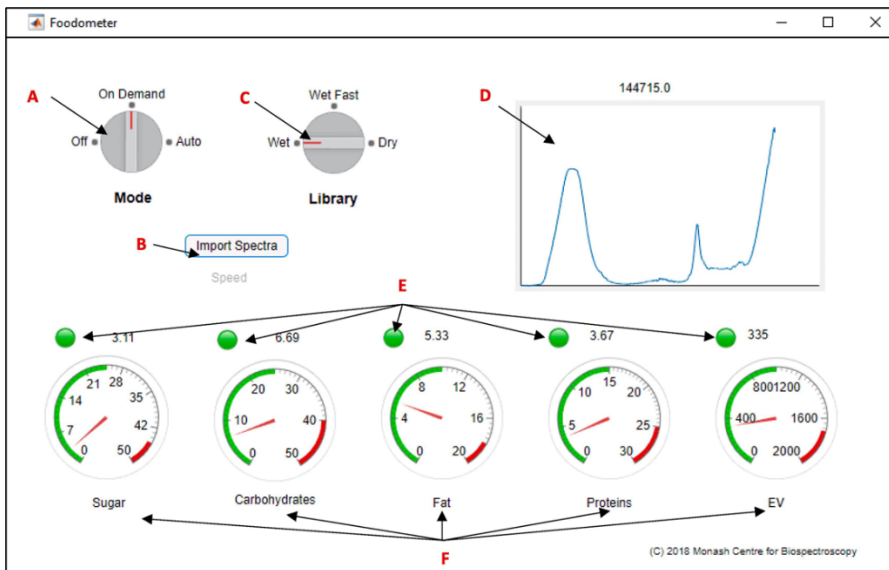


Figura 5. Captura de la App utilizada como ejemplo en el análisis multivariante.

4.3. Encuesta final

La Figura 6 ilustra los resultados de la encuesta final sobre la utilidad de MATLAB y Orange realizada a los alumnos. Se puede observar cómo los alumnos realizan una valoración muy positiva del programa interactivo Orange. Mas del 80% de los encuestados estaban de acuerdo o muy de acuerdo sobre su utilidad a la hora de entender y aplicar los métodos de tratamiento de datos. Además, el 64 % se ve preparado para usarlo en su trabajo de fin de máster. En contraste, la valoración del lenguaje de programación Matlab es mucho más negativa. A ningún estudiante le pareció útil a la hora de aplicarlo y sólo al 18% le ayudó a entender los conceptos quimiométricos en estudio.

Estos resultados indican claramente que los estudiantes prefieren aprender utilizando softwares intuitivos en comparación con complejos lenguajes de programación. Hay que entender que el uso de lenguajes de programación sólo se estudia brevemente en el primer curso del grado de Química. Tiene cierto sentido, por tanto, que los estudiantes tengan más dificultades a la hora de interactuar con Matlab. En este sentido, en las encuestas relativas a la comparación de software (Figura 7), los estudiantes muestran claramente que los problemas de MATLAB vienen de “Requerir muchas horas de aprendizaje” o “Ser difícil si no sabes programar”. Los estudiantes sí que entienden que el software Orange “Está limitado por las funciones que tiene” y “No tiene tantas opciones como Matlab”.

A pesar de que en general las opiniones sobre Orange son mucho más positivas, sorprendentemente, ante la pregunta “En un futuro si tuvieras más tiempo, ¿Sobre qué software te gustaría aprender más?”, sólo el 25 % de los alumnos indica Orange, otro 25% escoge Matlab y 50% elegiría los dos. En este sentido, los resultados indican que, pese a que no les fue útil, los alumnos comprenden el potencial del lenguaje de programación para el “Machine Learning” y están interesados en su aprendizaje. Parece ser que el poco tiempo de la asignatura es el factor limitante, como indican los alumnos al pedirles *feedback* (Ver Figura 8).

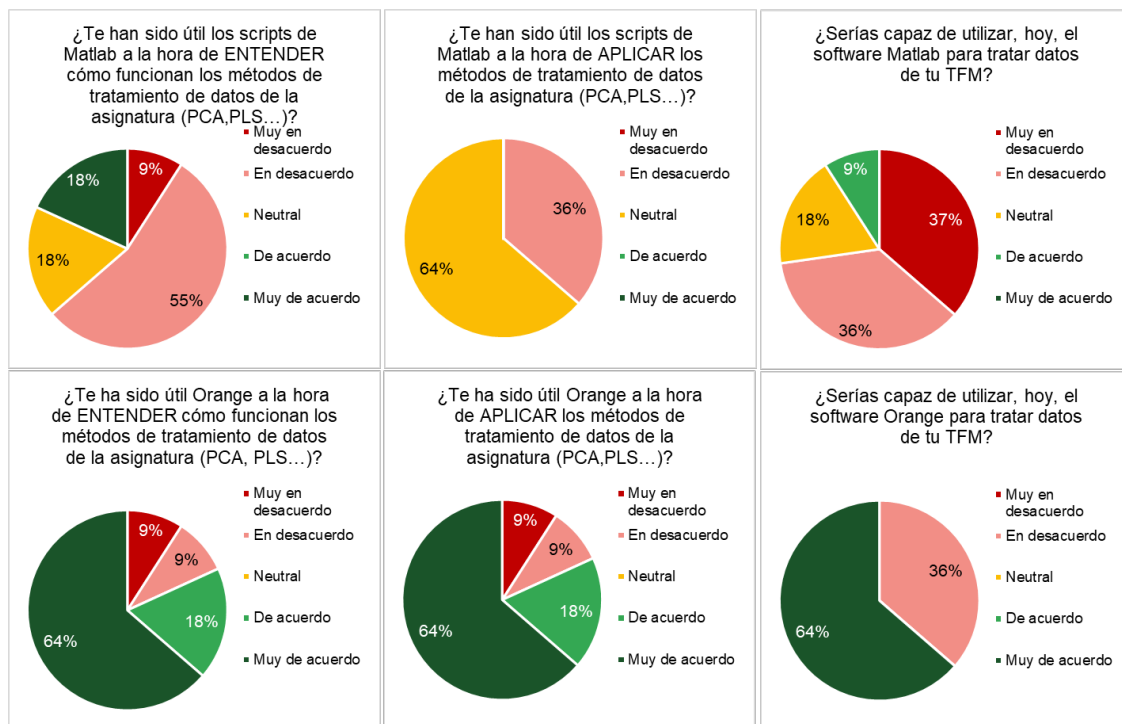


Figura 6. Resultados de la encuesta relativos a la utilidad y uso de Matlab y Orange.

La enseñanza del machine learning mediante propuestas prácticas e interactivas: la programación de aplicaciones en MATLAB para el análisis de datos espectrales de alimentos

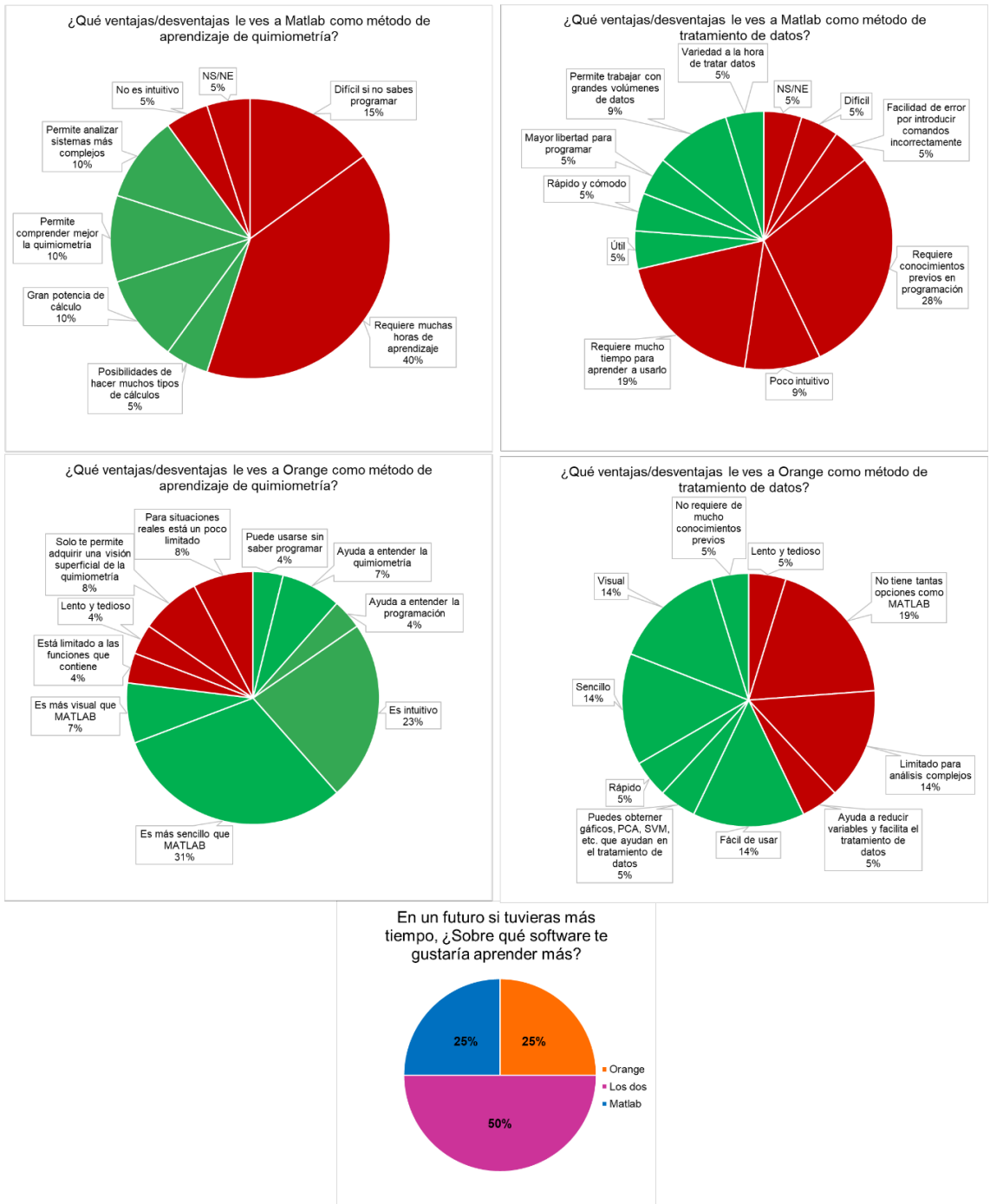


Figura 7. Resultados de la encuesta relativos a la comparación entre Matlab y Orange.

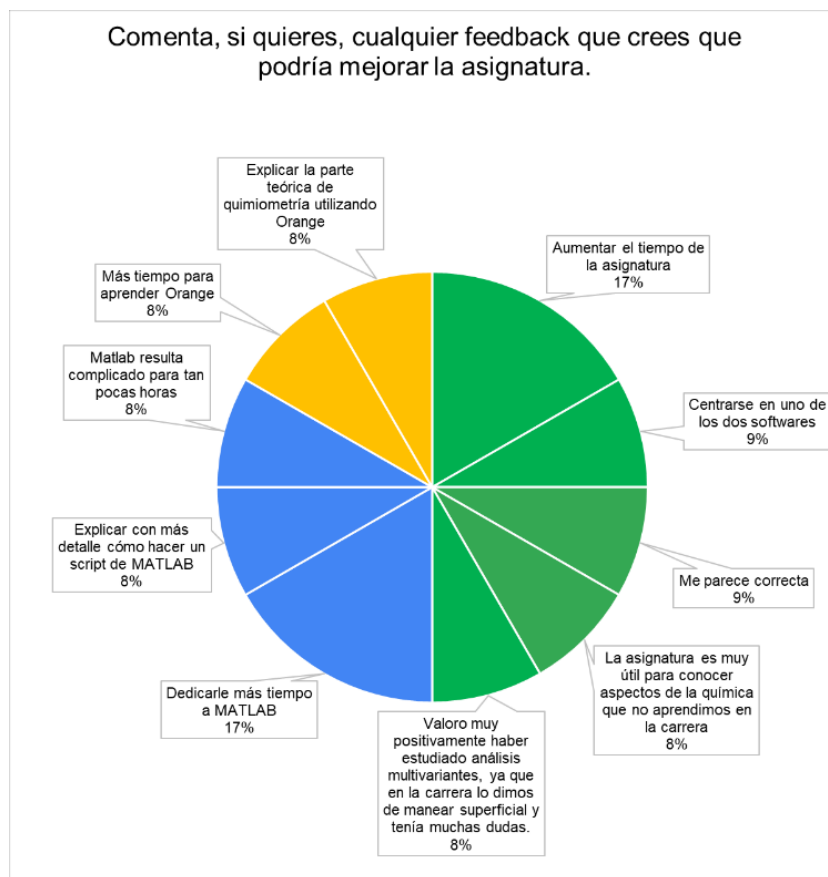


Figura 8. Feedback libre obtenido en la encuesta.

5. Conclusiones

Las encuestas iniciales indicaron que ningún alumno había realizado algún ejercicio práctico con softwares dedicados en el grado. La falta de ejercicios prácticos y la superficialidad con la que se tratan estos temas en el grado resulta en problemas conceptuales en conocimientos de estadística multivariante. Esta falta de conocimientos iniciales hacía difícil la adquisición de las competencias en quimiometría que figuran en la guía docente de la asignatura del máster. Los resultados de las encuestas indican que los ejercicios prácticos con softwares dedicados utilizando datos de IR de alimentos fueron beneficiosos para entender los modelos supervisados y no supervisados.

Respecto a la comparación entre programas intuitivos como Orange y lenguajes de programación como MATLAB, los resultados indican que para el formato de la asignatura Orange es más útil. El lenguaje de programación MATLAB, pese a ser atractivo para los alumnos, no es efectivo con las pocas horas dedicadas. Por ello, si de verdad se desea implementar la aplicación del *machine learning* en química, se debería apuntalar el uso de programación desde el inicio del grado. Cabe mencionar que este tipo de metodologías basadas en la incipiente inteligencia artificial no sólo son revolucionarias en química analítica, sino también en química computacional y bioquímica.

6. Referencias

- Debus, B., Parastar, H., Harrington, P., & Kirsanov, D. (2021). Deep learning in analytical chemistry. *TrAC Trends in Analytical Chemistry*, *145*, 116459. <https://doi.org/10.1016/j.trac.2021.116459>
- Delaney, M. F., & Warren, F. V. (1981). Teaching chemometrics: A course on application of mathematical techniques to chemistry. *Journal of Chemical Education*, *58*(8), 646. <https://doi.org/10.1021/ed058p646>
- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M., & Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, *14*, 2349–2353.
- Fangohr, H. (2004). A comparison of C, MATLAB, and python as teaching languages in engineering. In M. Bubak, G. DickVanAlbada, P. M. A. Sloot, & J. J. Dongarra (Eds.), *Computational Science—Iccs 2004, Proceedings* (Vol. 3039, pp. 1210–1217). Springer-Verlag Berlin. https://link.springer.com/content/pdf/10.1007/978-3-540-25944-2_157.pdf
- Holme, T. A. (2019). Can Today's Chemistry Curriculum Actually Produce Tomorrow's Adaptable Chemist? *Journal of Chemical Education*, *96*(4), 611–612. <https://doi.org/10.1021/acs.jchemed.9b00112>
- Howery, D. G., & Hirsch, R. F. (1983). Chemometrics in the chemistry curriculum. *Journal of Chemical Education*, *60*(8), 656. <https://doi.org/10.1021/ed060p656>
- Joshi, P. B. (2023). Navigating with chemometrics and machine learning in chemistry. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-023-10391-w>
- Perez-Guaita, D., Richardson, Z., Rajendra, A., Byrne, H. J., & Wood, B. (2021). From bench to worktop: Rapid evaluation of nutritional parameters in liquid foodstuffs by IR spectroscopy. *Food Chemistry*, *365*, 130442. <https://doi.org/10.1016/j.foodchem.2021.130442>
- St James, A. G., Hand, L., Mills, T., Song, L., Brunt, A. S. J., Bergstrom Mann, P. E., Worrall, A. F., Stewart, M. I., & Vallance, C. (2023). Exploring Machine Learning in Chemistry through the

Classification of Spectra: An Undergraduate Project. *Journal of Chemical Education*, 100(3), 1343–1350. <https://doi.org/10.1021/acs.jchemed.2c00682>

Tadayon, M. (2020). Python vs R vs Matlab for Machine Learning, Causal Inference, Signal Processing, and More. *The Startup*. <https://medium.com/swlh/python-vs-r-vs-matlab-for-machine-learning-causal-inference-signal-processing-and-more-b837a988c674>