

## DETERMINACIÓN DEL ORIGEN GEOGRÁFICO DE MIELES DE MENDOZA (ARGENTINA) MEDIANTE ANÁLISIS MULTIELEMENTAL Y ÁRBOLES DE DECISIÓN EN R

### Introducción

La miel es un alimento natural complejo producido por las abejas *Apis Mellifera* a partir del néctar, de las secreciones de plantas o de la excreción de insectos. Actualmente, la autenticidad de la miel se ha convertido en un tema de interés tanto para los productores como los consumidores, quienes reclaman información precisa sobre la procedencia de los alimentos y, por otro lado, los productores quieren evitar la competencia desleal y agregar valor a los productos que ofrecen.

La composición de la miel está relacionada con su origen botánico, que está estrechamente asociado a la zona geográfica en la que se origina, debido a las características del suelo y las condiciones climáticas estacionales, pero también está vinculada a factores como las especies de abejas, la temporada de producción y el almacenamiento, y las prácticas agrícolas en general. La composición elemental parece ser un buen indicador para los sistemas de clasificación geográfica, aunque esta sólo representa el 0.1-0.2% de la composición total de la miel, los minerales se transportan desde el suelo a las plantas y las flores a través del sistema radicular, pasan al néctar y finalmente a la miel. Según la FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación), Argentina está posicionada como el tercer mayor productor y exportador mundial de miel, exportando alrededor del 95% de su producción total, que es reconocida en todo el mundo por su alta calidad. Dentro de las regiones apícolas de Argentina, la provincia de Mendoza es considerada una de las más importantes del país.

En el software libre R, existe una variedad de algoritmos de clasificación y dentro de ellos, los árboles de decisión son un método de fácil interpretabilidad, pues nos dan un conjunto de reglas a partir de las cuales se pueden tomar decisiones. Además son algoritmos que no son demandantes en poder de cómputo comparado con procedimientos más sofisticados y dan buenos resultados de predicción para muchos tipos de datos.

El objetivo de este trabajo fue evaluar la capacidad de clasificación predictiva de la herramienta estadística multivariable "árboles de decisión" combinada con un análisis multielemental para diferenciar mieles de diferentes regiones apícolas de Mendoza (Argentina) de acuerdo con su región de origen. La originalidad del trabajo se enfoca en la aplicación de distintos algoritmos en R en este tipo de muestras de Mendoza para lograr la primera clasificación regional.

### Materiales y métodos

Se recolectaron 154 mieles de tres regiones apícolas de la provincia de Mendoza (Argentina): noreste (Lavalle, Rivadavia y Santa Rosa), Valle de Uco (San Carlos, Tunuyán y Tupungato) y sur (General Alvear, Malargüe y San Rafael). Las muestras se mineralizaron mediante una digestión ácida con HNO<sub>3</sub> concentrado en vaso abierto, aplicando un tratamiento térmico sobre una placa caliente durante 45 minutos a 100 °C. Las determinaciones analíticas se realizaron utilizando un espectrómetro de masas por plasma acoplado inductivamente (ICP-MS). Los isótopos detectados, en orden creciente del número de masa, fueron los siguientes: <sup>7</sup>Li, <sup>23</sup>Na, <sup>24</sup>Mg, <sup>27</sup>Al, <sup>44</sup>Ca, <sup>47</sup>Ti, <sup>51</sup>V, <sup>52</sup>Cr, <sup>55</sup>Mn, <sup>56</sup>Fe, <sup>59</sup>Co, <sup>60</sup>Ni, <sup>63</sup>Cu, <sup>66</sup>Zn, <sup>75</sup>As, <sup>78</sup>Se, <sup>85</sup>Rb, <sup>88</sup>Sr, <sup>95</sup>Mo, <sup>105</sup>Pd, <sup>107</sup>Ag, <sup>111</sup>Cd, <sup>118</sup>Sn, <sup>121</sup>Sb, <sup>201</sup>Hg, <sup>207</sup>Pb y <sup>208</sup>Pb.

Para el análisis estadístico, debido al desbalance en la cantidad de muestras entre las tres regiones y la importancia de la región del Valle de Uco como un poderoso polo apícola gracias a sus condiciones favorables para el desarrollo de la actividad, con ventajas que se refieren al clima y la cubierta vegetal, se procedió a clasificar miel de la región Valle de Uco de las demás regiones de Mendoza. Por lo tanto, todas las muestras que pertenecieron a la misma región fueron asignadas por el mismo código "V" para las mieles

del Valle de Uco y "O" para las mieles de las regiones del sur y noreste. Antes de aplicar los algoritmos de clasificación, se realizó un pre-procesamiento del conjunto de datos mediante la estandarización de la matriz. Todos los análisis estadísticos se calcularon utilizando el software R versión 3.3.0 con paquetes de árbol de decisión y caret. Se evaluaron cuatro algoritmos: recursive partitioning and regression trees (rpart), conditional inference tree (ctree), C5.0 y random forest (RF), utilizando la curva ROC (receiver operating characteristic), la cual es una representación gráfica que se utiliza para evaluar un sistema clasificador binario y seleccionar el modelo óptimo teniendo en cuenta AUC (área bajo la curva en su sigla en inglés). Para el desarrollo del proceso de clasificación, fue necesaria una separación del dataset en muestras de entrenamiento (70%) y de prueba (30%) y se aplicaron validaciones cruzadas de k-fold (CV) para la optimización del modelo.

## Resultados y Discusión

Luego del entrenamiento del dataset mediante 10-fold validación cruzada para construir los diferentes clasificadores, se seleccionó el modelo con el valor más grande de ROC para cada algoritmo de árbol de decisión. En los modelos de clasificación construidos la variable dependiente fue la región (V y O), y todos los elementos analizados se aplicaron como predictores. Se realizó una comparación del AUC para los diferentes clasificadores de árboles de decisión para la discriminación de origen de la miel en el conjunto de prueba. La Fig. 1 presenta la curva ROC específica para cada algoritmo evaluado para la clasificación de la miel de Mendoza. Los resultados mostraron un rendimiento similar para la predicción de la procedencia de la miel con un AUC superior a 0,9, que es un muy buen resultado que equilibra la sensibilidad y la especificidad. Los modelos de RF (AUC = 0.971) y ctree (AUC = 0.967) tienen el AUC más alto. Entre RF y ctree, este último puede interpretarse fácilmente como un solo árbol en lugar de un bosque (conjunto de árboles que es RF). C5.0 (AUC = 0.911) y rpart (AUC = 0.909) mostraron un poder de predicción similar. Esto se condice con los valores de accuracy obtenidos de 0.96 para RF y ctree, 0.93 para C5.0 y de 0.87 para el algoritmo rpart.

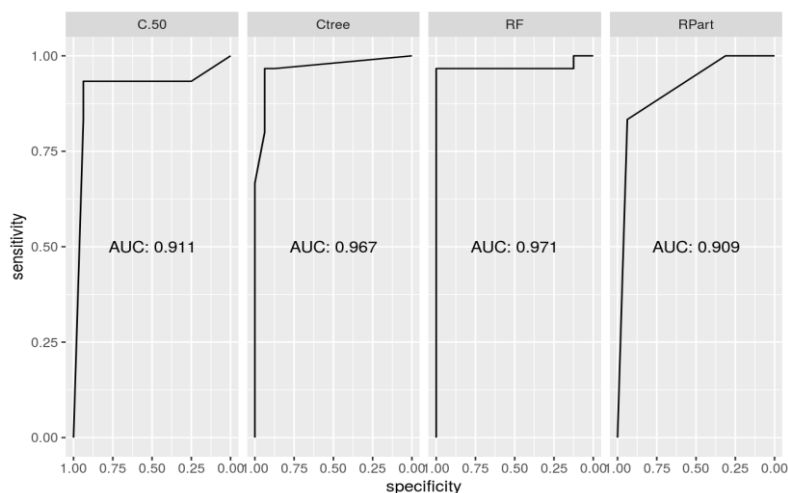


Fig. 1. Curvas ROC de los cuatro modelos probados.

## Conclusión

Con la aplicación de cuatro algoritmos de árbol de decisión, RF, ctree, rpart y C5.0, para discriminar el origen geográfico de miel de Mendoza, fue posible alcanzar una excelente clasificación del origen de este producto. Los resultados demostraron que los árboles de decisión son una forma simple y ágil de generar predictores altamente informativos de muestras de miel que se pueden usar para el control de la calidad y la verificación del origen geográfico.