



**UNIVERSIDAD
DEL AZUAY**

DEPARTAMENTO DE POSGRADOS

**Aplicación del aprendizaje supervisado para la predicción
del gusto de diversas moléculas**

Trabajo de titulación previo a la obtención del título de:

Magíster en Matemática Aplicada

Autor:

Mónica Fernanda Abril González

Director:

Cristian Rojas Villa

Cuenca-Ecuador

2023

DEDICATORIA

A Jorge, Ángel, Milanda, Alejandra y Andrea Abril.

A David Arízaga.

AGRADECIMIENTOS

Al Dr. Cristian Rojas, por todas sus enseñanzas y apoyo en el desarrollo de este trabajo.

Al Dr. Iván Mendoza y a la Dra. Verónica Pinos, por su apoyo constante durante la maestría.

ÍNDICE

| | |
|---|----|
| Resumen | 5 |
| Palabras clave | 5 |
| Abstract | 6 |
| 1. Introducción | 7 |
| 2. Materiales y métodos | 10 |
| 2.1 Generación de la base de datos | 10 |
| 2.2 Curado y Filtrado de la base de datos..... | 10 |
| 2.3 Cálculo de descriptores moleculares | 11 |
| 2.4 Desarrollo de modelos matemáticos de clasificación | 12 |
| 2.5 Evaluación de los modelos..... | 15 |
| 2.6 Validación de los modelos | 16 |
| 2.7 Programas quimioinformáticos | 17 |
| 3. Resultados y discusión | 17 |
| 3.1 Clasificación de compuestos dulces y amargos..... | 18 |
| 3.2 Clasificación de compuestos dulce-no dulce..... | 19 |
| 3.3 Clasificación de compuestos amargo-no amargo | 20 |
| 3.4 Clasificación de compuestos dulce-amargo-umami | 20 |
| 3.5 Predicción de los 5 gustos básicos | 21 |
| 4. Conclusiones | 23 |
| Referencias bibliográficas | 23 |

TABLA DE CUADROS

| | |
|---|----|
| Tabla 1 Modelos de clasificación para el gusto de moléculas desarrollados en los últimos cinco años y sus parámetros de evaluación NER (tasa de aciertos) en calibración y predicción..... | 9 |
| Tabla 2 Modelos QSAR basados en el aprendizaje automático de clasificación para la discriminación de los gustos dulce y amargo..... | 18 |
| Tabla 3 Modelos QSAR basados en el aprendizaje automático de clasificación para la discriminación de los gustos dulce y no dulce..... | 19 |
| Tabla 4 Modelos QSAR basados en el aprendizaje automático de clasificación para la discriminación de los gustos amargo y no amargo..... | 20 |
| Tabla 5 Modelos QSAR basados en el aprendizaje automático de clasificación para la discriminación de los gustos dulce, amargo y umami..... | 21 |
| Tabla 6. Modelos QSAR basados en el aprendizaje automático de clasificación para la discriminación de los cinco gustos básicos..... | 22 |
| Tabla 7. Índices primarios para el modelo de clasificación ECFPs-N3..... | 22 |

APLICACIÓN DEL APRENDIZAJE SUPERVISADO PARA LA PREDICCIÓN DEL GUSTO DE DIVERSAS MOLÉCULAS

Mónica Abril-González ¹, Cristian Rojas

Grupo de Investigación en Quimiometría y QSAR, Facultad de Ciencia y Tecnología, Universidad del Azuay, Av. 24 de Mayo 7-77 y Hernán Malo, Cuenca 010107, Ecuador.

¹ Correspondencia: mabrilma@es.uazuay.edu.ec

Resumen

La química del gusto es un tema de investigación importante en muchas disciplinas científicas, incluida la química de los alimentos. En los últimos años, se han propuesto varios modelos matemáticos basados en relaciones cuantitativas estructura-actividad (QSAR) para predecir el gusto de diversas moléculas. En este trabajo, una base de datos de 4116 estructuras moleculares fue curada y filtrada para desarrollar varios modelos de clasificación basados en el aprendizaje automático lineal y no lineal. Los gustos fueron representados por diversos descriptores moleculares, huellas dactilares y claves estructurales. Para el desarrollo de los modelos, la discriminación del gusto se consideró como clasificación binaria (dulce-amargo, dulce-no-dulce y amargo-no-amargo) y multiclase (dulce-amargo-umami y cinco sabores básicos). Para la validación de los modelos, los conjuntos de datos se dividieron en conjuntos de entrenamiento y predicción en una proporción de 70:30, manteniendo las proporciones de las clases en ambos grupos. Para la clasificación binaria, los métodos de los bosques aleatorios y boosting adaptativo con los descriptores moleculares fueron los clasificadores óptimos; mientras que el clasificador de los N -vecinos más cercanos con las huellas dactilares de conectividad ampliada fue el ideal para la clasificación multigusto.

Palabras clave

Química del gusto; Aprendizaje automático; Modelos de clasificación; QSAR.

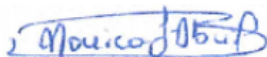
Abstract

The chemistry of taste is an important research topic in many scientific disciplines including food chemistry. In recent years, several mathematical models based on quantitative structure-activity relationships (QSARs) have been proposed to predict the taste of diverse molecules. In this work, a curated database of 4116 molecular tastants was filtered to develop several models based on linear and non-linear machine learning classifiers. Tastants were represented by diverse molecular descriptors, fingerprints and structural keys. For the development of models, taste discrimination was considered as either binary classifiers (sweet-bitter, sweet-non-sweet and bitter-non-bitter) or multiclass (sweet-bitter-umami and five basic tastes). For validation purposes, datasets were split into training and test sets in a proportion of 70:30 maintaining class proportions in both sets. For binary classification, the random forests and adaptive boosting methods with the molecular descriptors were the optimal classifiers; while the N-nearest neighbors classifier with the extended connectivity fingerprints was the ideal classifier for multitaste classification.

Keywords

Taste Chemistry; Machine Learning; classification models; QSAR.

Translated by:



Mónica Abril

