

¿CÓMO RECONOCER HONGOS VENENOSOS O COMESTIBLES?

Sanabria Martínez D. M.
González Castillo F.I.
Llanos Morales A.



INTRODUCCIÓN

Teniendo en cuenta que en la actualidad sólo se han clasificado un 10% de los 10 millones de especies de hongos en el mundo y que estos son de gran utilidad en áreas de la medicina, gastronomía y biología ha sido necesario el desarrollo de programas que realicen una clasificación efectiva de los hongos como venenosos o comestibles en base a sus características físicas.

La correcta clasificación de hongos es importante debido que son utilizados como alimento en diferentes circunstancias y podría significar una elección de vida o muerte para dicha persona, además, como se mencionó anteriormente la clasificación es utilizada en diferentes áreas de la ciencia, como por ejemplo, en platillos de cocina, en la realización de fármacos o en investigaciones científicas.

Por lo que utilizaremos análisis estadístico y las técnicas de árbol de decisión y visualización para la realización del proyecto.

OBJETIVOS

Saber cuáles son las principales características que determinan si un hongo es venenoso o comestible.

Así como poder realizar una clasificación completa y efectiva de la totalidad de hongos en la base de datos en el menor tiempo posible.

RECURSOS



Librerías utilizadas:

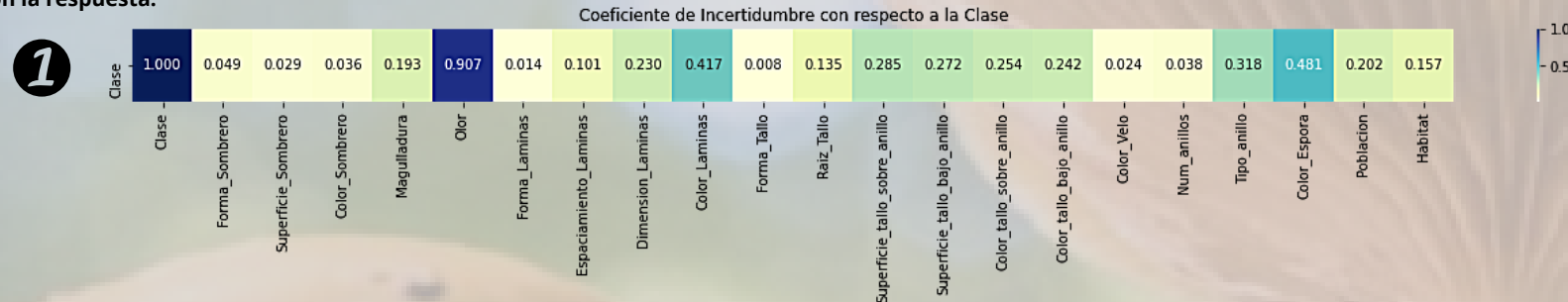
- Pandas
- Seaborn
- Matplotlib
- Math
- Scipy.stats
- Sklearn

Base de datos

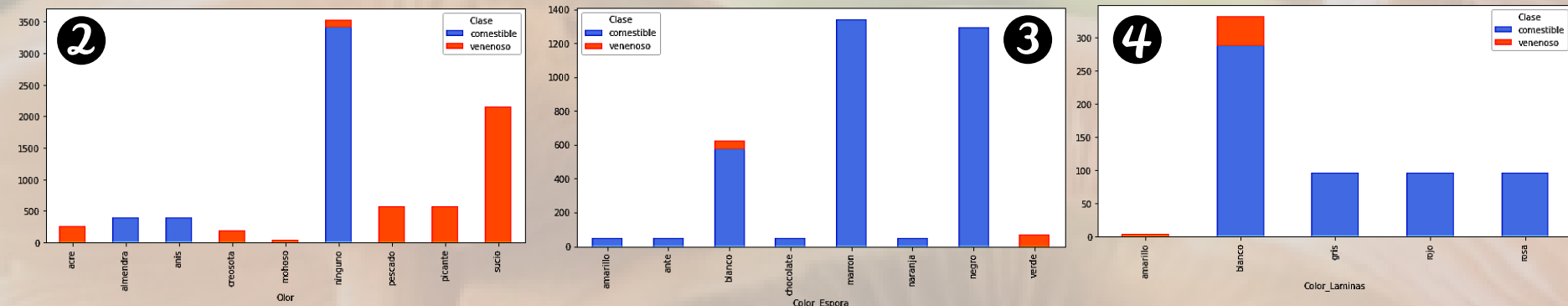


RESULTADOS

Con ayuda del Coeficiente de incertidumbre (gráfica 1) podemos observar que las características de Olor, Color de Láminas y Color de Esporas son las que tienen una mayor relación con la respuesta:



Después de haber encontrado las categorías que mas influyen, utilizamos la técnica de Visualización con la ayuda de tablas de contingencia, graficas y matrices de correlacion, en ellas pudimos observar que:



Revisando la categoría de olor, llegamos a las siguientes conclusiones:

- Si nuestro hongo tiene el olor acre, creosota, mohoso, pescado, picante o sucio son venenosos.
- Si nuestro hongo tiene el olor a almendra o anís se trata de uno comestible.
- En caso de que nuestro hongo no presente olor, entonces, no podremos dar una respuesta con esta información.

Si nuestro hongo no presentó ningún olor, chequearemos el color de su espора:

- Si su color es amarillo, ante, chocolate, marrón, naranja o negro es comestible.
- En cambio, si su color de espора es verde se trata de un hongo venenoso.
- En caso de que sea blanco, entonces, no podremos dar una respuesta con esta información.

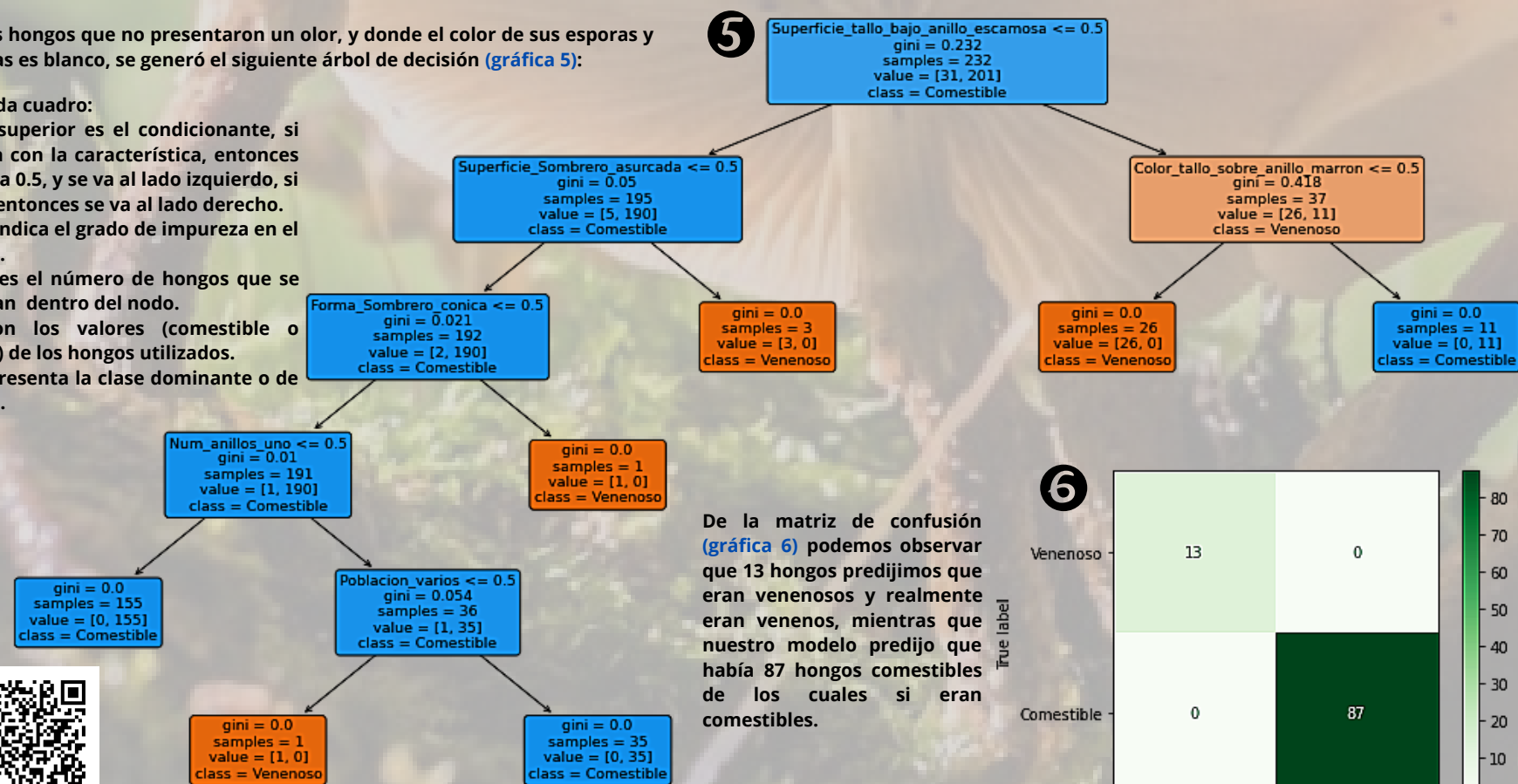
Si nuestro hongo no presentó ningún olor y además las esporas son de color blanco, chequearemos el color de sus láminas:

- Si su color son gris, rojo o rosa son hongos comestibles.
- Si el color de sus laminas son amarillo se trata de un hongo venenoso.
- En caso de que sea blanco, entonces, no podremos dar una respuesta con esta información.

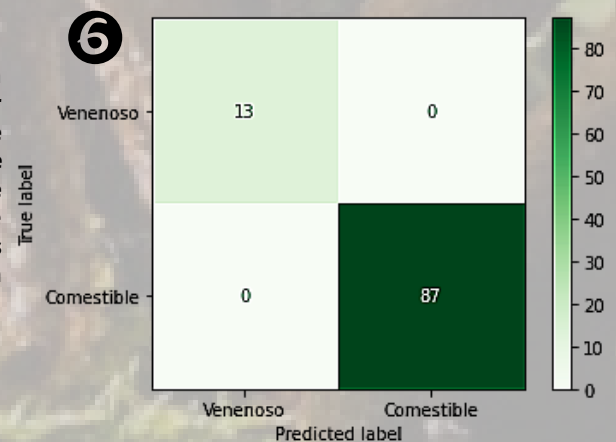
Para aquellos hongos que no presentaron un olor, y donde el color de sus esporas y de sus láminas es blanco, se generó el siguiente árbol de decisión (gráfica 5):

Dentro de cada cuadro:

- La parte superior es el condicionante, si no cuenta con la característica, entonces es menor a 0.5, y se va al lado izquierdo, si la posee, entonces se va al lado derecho.
- Gini: nos indica el grado de impureza en el segmento.
- Samples: es el número de hongos que se encuentran dentro del nodo.
- Value: son los valores (comestible o venenoso) de los hongos utilizados.
- Class: representa la clase dominante o de respuesta.



De la matriz de confusión (gráfica 6) podemos observar que 13 hongos predijimos que eran venenosos y realmente eran venenos, mientras que nuestro modelo predijo que había 87 hongos comestibles de los cuales si eran comestibles.



METODOLOGÍA

La base de datos utilizada está conformada por 23 características y registro de 8124 hongos, con sólo atributos cualitativos, principalmente de su forma y color.

Para poder trabajar se validaron datos nulos, los cuales no fueron encontrados por lo que no fue necesario eliminar los datos en este punto, en cambio se eliminó la columna de *Tipo_Velo*, ya que no representaba relevancia. Para un mejor entendimiento se tradujo cada variable al español.

Se genera una prueba preliminar estadística con el coeficiente de incertidumbre (gráfica 1) determinando las características más relevantes.

A continuación se realizó una clasificación utilizando las gráficas 2, 3 y 4, además de tablas de frecuencias con la característica que más influía, después se creó un nuevo DataFrame para aquellos hongos que aún no habían sido clasificados. Este proceso se repitió para los análisis de las tres características principales.

Debido a que todas nuestras variables eran categóricas se cambiaron a variables cuantitativas, por medio del código de variables dummies.

Después se dividió la base de datos resultante (332 datos) de los procesos anteriores en 70% para los datos de entrenamiento (232 datos) y 30% para los datos de prueba (100 datos). Con los datos de entrenamiento se aplicó el algoritmo de árbol de decisión y con los datos de prueba se generó la matriz de confusión de los resultados.

CONCLUSIONES

Gracias al análisis que realizamos a la base de datos pudimos identificar en la gráfica 1 las principales características que determinan si un hongo es venenoso o comestible, estas en orden de relevancia son Olor, Color_Espora y Color_Laminas.

También podemos decir que cumplimos con nuestro objetivo secundario, ya que con ayuda del árbol de decisión (gráfica 5) para las características restantes se logró clasificar en su totalidad a los hongos en la base de datos, y que, al ser un esquema de fácil entendimiento consigue que la clasificación se realice en un tiempo menor.

En un futuro se podrían utilizar los resultados obtenidos para la creación de una aplicación en la cual por medio de una fotografía del hongo el algoritmo identifique estas características determinantes y le responda al usuario si dicho hongo es venenoso o comestible.