



Laboratorio 1 - Análisis Estadístico

Integrantes: William Martínez

Felipe Muñoz

Curso: Análisis de Datos

Sección A-1

Profesor: Max Chacón Pacheco

21 de Septiembre de 2022

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	1
2. Descripción del problema	2
2.1. El hongo, un organismo único	2
2.2. Descripción de la base de datos	3
2.3. Descripción de clases y variables	3
3. Análisis estadístico e Inferencial	10
3.1. Prueba de independencia Chi-cuadrado	11
3.2. Coeficiente V de Cramer	12
3.3. Gráficas de frecuencias con respecto a tablas de contingencias	13
4. Conclusión	21
Bibliografía	23

1. Introducción

Los hongos suelen pasar bastantes desapercibidos en la vida de las personas y menudo son vistos como un simple producto para ser consumido o sencillamente algo venenoso. La mayoría de los seres humanos no son conscientes de la gran importancia que poseen los hongos para la conservación de la diversidad los bosques y su capacidad de confrontar el cambio climático amortiguando los impactos provocados.

”La seta es al hongo lo que el fruto es al árbol”. El hongo a través de la seta es capaz de esparcirse mediante esporas almacenadas y expulsadas bajo el sombrero. Por otro lado, aquel aspecto que no se puede ver de la seta (el hongo que se encuentra bajo la superficie) permite conectar todo el ecosistema por medio de una red compleja denominada ”Wood Wide Web”, es español ”la internet del bosque”(Magaña, 2019).

Los hongos suelen ser clasificados en comestibles, venenosos, medicinales y alucinógenos. En este informe se estudiará principalmente los dos primeros mediante sus principales características, ya sea, forma de sombrero, textura del tronco, número de anillo, color de esporas, entre otras. Es gracias a un análisis estadístico descriptivo e inferencial que se llevarán a cabo conclusiones con respecto a las características de los hongos y cuan determinantes son estas a la hora de clasificarlos. La base de datos es recopilada mediante la página web *UCI-Machine Learning Repository* y el análisis se llevará por medio del Software RStudio.

1.1. Objetivos

1. Comprender las distintas características asociadas a los hongos a través de la base de datos recopilada.
2. Llevar a cabo tanto un análisis estadístico descriptivo como inferencial de la información obtenida de la base de datos perteneciente a la setas.
3. Interpretar los resultados con un razonamiento objetivo y coherente mediante el uso correcto de modelos estadísticos.

2. Descripción del problema

2.1. El hongo, un organismo único

Los hongos han existido hace más de mil millones de años en el planeta tierra, y el hecho de que aún existan da a conocer cuán esenciales son para la vida del ecosistema. Los hongos antiguamente se han considerado como parte del reino vegetal, sin embargo con el paso del tiempo científicos se han interiorizado más en el tema y han concluido que estos organismos no son ni plantas ni animales, más bien pertenecen a un reino totalmente distinto, "El Reino Funji". Los hongos se destacan principalmente por alimentarse a través de la "absorción", a diferencia de las plantas que obtienen sus nutrientes mediante el proceso de fotosíntesis y los animales por medio de la ingestión de alimentos específicos (Furci and Magnasco, 2017a).

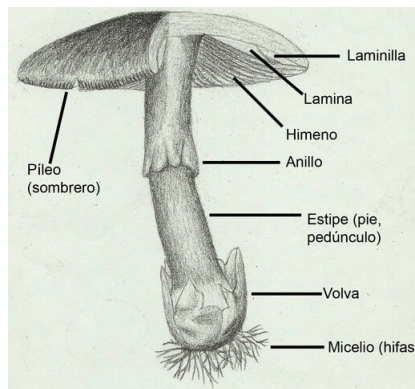


Figura 1: Partes del cuerpo fructífero del hongo.

No es necesario manifestar que los hongos son organismos esenciales en la degradación de materia orgánica y que causan un gran misterio con el solo hecho de observarlos. Estos están presentes alrededor de todo el mundo, ya sea en aguas, superficies terrestres, bosques, en el aire e incluso en algunos cultivos. Hoy en día su existencia permite comercializarlos en ámbitos alimenticios y medicinales. La subsistencia de los hongos ha provocado que existan una gran variedad de estos, partiendo desde la diversidad de sus tamaños, hasta los distintos colores y formas que pueden llegar a tener. La mejor época para encontrar hongos es en otoño, estos suelen desarrollarse en lugares húmedos y en ausencia de la luz, ya que no es la energía vital para su reproducción. Los hongos se reproducen por medio de esporas que se forman en el himenio del cuerpo de la seta (véase Figura 1). Las esporas pueden llegar a ser un buen identificador para clasificar hongos, ya que varían en su color y forma (elipsoides, cilíndricas, lisas, entre otras). Los hongos se encuentran en las raíces de los árboles, arbustos e

incluso en la gran mayoría de plantas, es allí donde ocurre la simbiosis (Furci and Magnasco, 2017b), proceso en el cual *"el hongo obtiene un lugar donde vivir y obtiene carbohidratos, mientras que la planta obtiene una mayor capacidad de recibir nutrimentos del suelo y del agua"* (Torres, 2013).

El estudio de la anatomía de las setas, el tipo de esporas que expulsa o la forma en que se reproduce, permite clasificar a los hongos en varios tipos, desde comestibles, hasta tóxicos, alucinógenos e incluso mortales. Es debido a esto que es de suma importancia llevar a cabo un estudio sobre como se relacionan las características anatómicas de un hongo con su clasificación, para este caso, "comestibles" y "venenosos".

2.2. Descripción de la base de datos

La base de datos a analizar reúne datos de hongos en términos de sus características físicas, clasificándolos de acuerdo a su comestibilidad. Esta base de datos fue donada el 27 de abril de 1987 por Jeff Schlimmer al UCI (*Machine Learning Repository*) desde donde se extrajo para examinarla. Su origen se encuentra en el registro de hongos extraídos de la guía de campo para hongos norteamericanos de la Sociedad Audubon del año 1981. En específico, la base de datos describe 23 hipotéticas especies de hongos sin branquias del tipo *Agaricus* y *Lepiota*. Estas pueden ser clasificadas como definitivamente comestibles, definitivamente venenosas y de comestibilidad desconocida (a esta última clasificación es tomada como venenosa por la base de datos). En términos técnicos la base de datos cuenta con 8124 observaciones, posee 22 variables del tipo categóricas las cuales pueden contener valores perdidos (UCI, sf) y dos grandes clases (comestible y venenoso).

2.3. Descripción de clases y variables

Las clases que se analizarán, como se mencionó en el apartado anterior, corresponden a aquellas que dan a conocer aquellos hongos que son comestibles o venenosos. En concreto, las clases son edible (comestible), representada en la base de datos por el valor "e"; y poisoned (venenoso), representada en la base de datos por el valor "p".

Dado que los atributos de la base de datos corresponden a variables categóricas, es justo manifestar que estos representan cualidades físicas de los hongos, por lo tanto las magnitudes cuantitativas no son reflejadas para este estudio. En razón de lo anterior, es que se crea una tabla que describe cada una de las características que tiene un seta y los respectivos valores que puede optar. (ver Figura 2, 3, 4, 5, 6). La información encontrada pertenece

a aquella recopilada de diversas páginas web, como por ejemplo (Tullii, 2021) (Fondation Miron-Royer Inc, 2018) y (Osorio, 2021)

Variables	Descripción	Atributos que puede tomar
1. cap-shape: forma de sombrero.	Parte superior de la seta, que se encuentra sobre el tronco(stalk). Su función biológica consiste en resguardar el himenio y facilitar así el correcto desarrollo de las esporas.	b=bell (campana) c=conical (cono triangular) x=convex (convexa) k=knobbed (also bump or knob: bulto) s=sunken (hundido)
2. cap-surface: superficie de sombrero.	Textura de la superficie del píleo (sombrero de la seta). Facilita la identificación del tipo de hongo.	f=fibrous (fibrosa) g=grooves (surcos) y=scaly (escamosa) s=smooth (suave)
3. cap-color: color del sombrero.	Corresponde al color que puede tomar el sombrero del cuerpo fructífero del hongo.	n=brown (café) b=buff (ante) c=cinnamon (canela) g=gray (gris) r=green (verde) p=pink (rosado) u=purple (morado) e=red (rojo) w=white (blanco) y=yellow (amarillo)
4. bruises: moretones.	Para ser más específicos, los moretones azules pueden indicar si un hongo contiene o no psilocibina y psilocina, los principales compuestos químicos responsables de los efectos psicoactivos. (psilocibina: Algunos tipos de setas contienen psilocibina y psilocina, unas sustancias que pueden causar alucinaciones. Cuando se usan a dosis altas, estas setas pueden tener efectos similares a los de la droga LSD).	bruises=t or no=f (¿tiene o no tiene moretones?)

Figura 2: Descripción de los atributos y sus respectivos valores.

5. odor: olor.	Los hongos emanan una gran variedad de olores que son efectuados ya sea por su metabolismo intrínseco o por estrategias reproductivas.	a=almond (almendra) l=anise (anís) c=creosote (olor fuerte y ahumado, camino recién pavimentado de asfalto) y=fishy (olor a pescado-Lactifluus volemus) f=foul (olor fétido) m=musty (olor rancio; mohoso) n=none (sin olor) p=pungent (olor acre-ajo,fósforo) s=spicy (picante)
6. gill-attachment: accesorio branquial.	Cada seta cuenta en el himenio con cientos de laminillas, rodeadas todas ellas por innumerables basidios (estructura microscópica productora de esporas).	a=attached (adjunto-unido) d=descending (descendiendo) f=free (libre) n=notched (mellado: carece de)
7. gill-spacing: Espacio entre branquias	Corresponde al espacio que se genera entre las laminillas ubicadas en el himenio. dependiendo del tipo, estas pueden actuar como segunda capa protectora.	c=close (cerca) w=crowded (apretado) d=distant (distante)

Figura 3: Continuación tabla - Descripción de los atributos y sus respectivos valores.

8. gill-size: Tamaño de branquias	Las laminillas o branquias pueden tener diferentes tamaños, desde muy pequeñas hasta bastante amplias.	b=broad (amplio) n=narrow (angosto)
9. gill-color: Color de branquias	Las laminillas pueden tener diferentes colores, lo que ayuda también a la clasificación de los hongos.	k=black (negro) n=brown (café) b=buff (ante) h=chocolate g=gray (gris) r=green (verde) o=orange (anaranjado) p=pink (rosado) u=purple (morado) e=red (rojo) w=white (blanco) y=yellow (amarillo)
10. stalk-shape: Forma de pie o tronco.	Su función consiste en elevar el sombrero y el himenio sobre la maleza, favoreciendo así la dispersión de las esporas en el ambiente.	e=enlarging (dilatado, extender) t=tapering (estrechado)
11. stalk-root: Raíz de tallo o tronco	Corresponde a la estructura que posee la raíz del tallo de la seta. Su forma puede indicar información relevante en un estudio determinado.	b=bulbous (abultada;bulboso) c=club (raíz hinchada y distorsionada; enfermedad en las plantas) u=cup (copa) e=equal (igual) z=rhizomorphs (rizomorfo;forma de raíz) r=rooted (enraizado) missing=?

Figura 4: Continuación tabla - Descripción de los atributos y sus respectivos valores.

12. stalk-surface-above-ring: Superficie del tronco sobre el anillo	Esta variable hace referencia al tipo de superficie del tallo por sobre el anillo. El tipo de superficie puede dar cuenta de las características que posea el hongo. Por ejemplo una superficie suave es prueba de la falta de ornamentación en aquella parte del tallo, mientras que una escamosa puede ser resultado de la acumulación de fibras que se unen unas a otras, formando "parches".	f=fibrous (fibroso) y=scaly (escamosa) k=silky (sedosa) s=smooth (suave)
13. stalk-surface-below-ring: Superficie del tallo bajo el anillo	Esta variable describe la superficie de la parte inferior del tallo de los hongos. Se sabe que la superficie del tronco bajo el anillo puede diferir de la que se encuentra sobre este, denotando la naturaleza del hongo, por lo que es necesario estudiarla.	f=fibrous (fibroso) y=scaly (escamosa) k=silky (sedosa) s=smooth (suave)
14. stalk-color-above-ring: Color del tallo sobre el anillo	Corresponde al color del tronco o tallo que se encuentra sobre el anillo. Su color puede ir variando dependiendo del tipo de hongo.	n=brown (café) b=buff (ante) c=cinnamon g=gray (gris) o=orange (anaranjado) p=pink (rosado) e=red (rojo) w=white (blanco) y=yellow (amarillo)

Figura 5: Continuación tabla - Descripción de los atributos y sus respectivos valores.

Variables	Descripción	Atributos que puede tomar
15.stalk-color-below-ring: Color del tallo bajo el anillo	Corresponde al color del tronco o tallo que se encuentra bajo el anillo. Su color puede ir variando dependiendo del tipo de hongo.	n=brown (café) b=buff (ante) c=cinnamon g=gray (gris) o=orange (anaranjado) p=pink (rosado) e=red (rojo) w=white (blanco) y=yellow (amarillo)
16.veil-type: Tipo de velo.	Esta variable define el tipo de velo que poseen los hongos. En su concepción los hongos se encuentran recubiertos por una estructura denominada velo universal. Mientras el hongo crece este velo comienza a desgarrarse. Por otra parte se encuentra el velo parcial, que cubre y protege al himenio (ejemplares jóvenes). Cuando el sombrero se abre, este se desprende dando origen al anillo.	p=partial (parcial) u=universal
17.veil color: Color de velo	Corresponde al color del velo que recubre a los hongos.	n=brown (café) o=orange (anaranjado) w=white (blanco) y=yellow (amarillo)
18.ring-number: Número de anillos	Se enumera la cantidad de anillos por cada observación de hongos. Los anillos se forman a la mitad del tronco una vez que este ha madurado y se ha roto el velo parcial. Su presencia y cantidad por cada hongo puede ser variable.	n=none (ninguno) o=one (uno) t=two (dos)

Figura 6: Continuación tabla - Descripción de los atributos y sus respectivos valores.

Variables	Descripción	Atributos que puede tomar
19. ring-type: Tipo de anillo	La variable indica el tipo de anillo de acuerdo a su estructura y forma. El tipo de anillo se puede clasificar de acuerdo al esquema de la figura 2.	c=cobwebby (fino) e=evanescent (evanescente) f=flaring (resplandeciente) l=large (largo/amplio) n=none (ninguno) p=pendant (colgante) s=sheathing (cubierto) z=zone (zona)
20. spore-print-color: Impresión de color de las esporas	La variable identifica el color que aparece posterior a la impresión del color de las esporas. Esto consiste en el resultado del depósito del polvo que se obtiene al dejar que estas caigan en conjunto sobre una superficie. Gracias a esto es posible dilucidar el color de las esporas.	k=black (negro) n=brown (café) b=buff (ante) h=chocolate r=green (verde) o=orange (anaranjado) u=purple (morado) w=white (blanco) y=yellow (amarillo)
21. population: Población	Esta variable hace referencia al tamaño y distribución de la población de los diferentes tipos de hongos en estudio.	a=abundant (abundante) c=clustered (agrupado) n=numerous (numeroso) s=scattered (disperso) v=several (varias) y=solitary (solitaria)
22. habitat: Hábitat	Esta variable señala el hábitat del hongo. Los hongos pueden nacer y crecer prácticamente en cualquier tipo de hábitat, ya sean zonas abiertas, húmedas, ácidas o básicas, arbóreas o arbustivas, etc. Además, pueden haber especies de hongos que sean capaces de colonizar diversos hábitats.	g=grasses (pastos) l=leaves (hojas) m=meadows (praderas) p=paths (senderos) u=urban (urbano) w=waste (desperdicios) d=woods (madera)

Figura 7: Continuación tabla - Descripción de los atributos y sus respectivos valores.

3. Análisis estadístico e Inferencial

Tomando en consideración el problema a resolver, y que las variables a analizar son del tipo categóricas, lo primero que se lleva a cabo, sin contar la lectura del archivo en R, es la organización de todas las variables en tablas de contingencia para dos variables. En este caso se estaría contando la aparición de cada valor de las variables con respecto a las clases en estudio, junto a una variable extra en donde se presenta la diferencia entre los valores encontrados. Debido a la cantidad de tablas resultantes, se ejemplifica el resultado mostrado en la Figura 8, en donde se destacan aquellas que presentan las mayores diferencias encontradas.

\$odor			
	e	p	diferencia
a	400	0	400
c	0	192	192
f	0	2160	2160
l	400	0	400
m	0	36	36
n	3408	120	3288
p	0	256	256
s	0	576	576
y	0	576	576
\$gill_size			
	e	p	diferencia
b	3920	1692	2228
n	288	2224	1936
\$bruises			
	e	p	diferencia
f	1456	3292	1836
t	2752	624	2128

Figura 8: Tablas de contingencia.

De los resultados obtenidos es destacable el hecho de que ciertas variables que poseen múltiples niveles pueden pasar a ser consideradas como dicotómicas, pues algunos de sus valores no coincidían ni con hongos venenosos ni con hongos comestibles. En específico, este fue el caso para las variables "gill-attachment" y "gill-spacing", para las cuales los atributos 'descending' y 'notched' en el caso de la primera variable; y 'distant' en el caso de la segunda fueron descartadas por no ser representativas de las clases. Otra situación que vale la pena recalcar es la ocurrida con la variable dicotómica 'veil-type'. Esto último a causa del hecho de que su atributo "universal" no es parte de la tabla de contingencia, lo cual lleva a pensar que todos los hongos en estudio solo poseen su velo parcial. Es por esto que se toma la decisión de descartar esta variable del estudio por medio de la instrucción 'NULL', por la nula información que podría aportar.

3.1. Prueba de independencia Chi-cuadrado

Si bien la información entregada por las tablas puede sugerir la implicancia que poseen ciertos valores de las variables sobre si un hongo es comestible o no, se hace necesario estudiar si en efecto dicha relación existe, pues, si no es el caso, entonces se deberá proceder a eliminar las variables que no se encuentren relacionadas al valor de las clases, puesto que no estarían aportando información relevante para el análisis. Con el fin de comprobar lo anterior, se procederá en primera instancia, a descartar aquellas variables que mantengan un comportamiento independiente. Para lograr esto se efectúa una prueba chi-cuadrado de independencia, debido a que esta "permite determinar si dos variables categóricas, de una misma población, son estadísticamente independientes o si, por el contrario, están relacionadas" (Jara and Köhler, 2021a). Esto último por medio de la función *chisq.test()* y de la función *lapply()*, con las cuales se recorren cada tabla de contingencia y se obtiene el valor "p" de la prueba que determinará si, por un lado, las variables son independientes (hipótesis nula), o se encuentran relacionadas (hipótesis alternativa). Es posible utilizar esta prueba porque, por un lado, se cumple el supuesto de independencia (los datos representan a menos del 10 % de la población mundial de hongos), y por otro la cantidad de observaciones esperadas por cada uno de los grupos es superior a 5 observaciones (esto se comprueba al momento de proceder con la prueba y conseguir los resultados). Si hubieran tablas de contingencia que no cumplieren con la condición anterior, será necesario aplicar una prueba exacta de Fisher, junto a aquellas tablas en que ambas variables son dicotómicas ("bruises", "gill-attachment", "gill-spacing", "gill-size" y "stalk-root"). La prueba exacta de Fisher funciona como alternativa a la prueba chi-cuadrado de independencia, incorporando las mismas hipótesis nulas y alternativas, pero siendo realizada por medio de la función "fisher.test" (Jara and Köhler, 2021b). Luego de revisar manualmente cada uno de los valores esperados obtenidos de las pruebas, se llegó a la conclusión presentada en la tabla de la Figura 9

Como resultado de aplicar la prueba de independencia se obtuvo que todas aquellas variables que cumplieron con la segunda condición obtuvieron un valor p menor a $2.2e-16$, por ende, con cualquier nivel de significación (para este caso un $\alpha=0.01$), se puede concluir que con 99 % de confianza es rechazada la hipótesis nula en favor de la alternativa, por lo tanto todas estas variables poseen relación con las clases en estudio. Por otro lado, luego de aplicar la prueba de Fisher con un nivel de significación alto ($\alpha = 0.01$) para las variables dicotómicas y aquellas que no cumplieron con la condición de los valores esperados (tablas con índice 1, 2, 14 y 16) se llegó a la misma conclusión previa, es decir las variables en estudio presentan algún tipo de relación con las clases (las tablas 1, 2 y 14 no es posible

Cumple	No cumple
cap_color	Cap_shape
bruises	Cap_surface
odor	Stalk_color_above_ring
gill_attachment	Veil_color
gill_spacing	
gill_size	
gill_color	
stalk_shape	
stalk_root	
stalk_surface_above_ring	
stalk_surface_below_ring	
stalk_color_below_ring	
ring_number	
ring_type	
spore_print	
population	
habitat	

Figura 9: Cumplimiento valores esperados por cada variable.

calcular valor directamente, por ende se simula el p-valor con 10000 réplicas obteniéndose un valor $p=9.999e-05$). Es así como, por el momento, no se deben seguir descartando variables, puesto que no son independientes de las clases.

3.2. Coeficiente V de Cramer

Antes de introducir un análisis profundo en lo que respecta al grado de relación que existe entre dos variables categóricas, es esencial describir que es el coeficiente "V de Cramer". Este coeficiente, trabaja como una medida estadística basada en Chi-cuadrado que tiene como objetivo calcular el grado de asociación que existe entre dos variables categóricas. De acuerdo a lo expresado, dicho coeficiente puede tomar valores entre 0 y 1. Para valores cercanos a 0, se concluye que el grado de asociación entre ambas variables es débil, mientras que para valores cercanos a 1 la asociación será fuerte.

Ya con el conocimiento del coeficiente de "V de Cramer", se procede a efectuar un análisis descriptivo a las variables recopiladas de la base de datos con el fin de discernir que atributos son los que poseen mayor relevancia al momento de realizar el estudio. Para poder llevar a cabo dicho procedimiento, se hace uso de la función "cramerV" de la biblioteca "rcompanion" en RStudio. Sin más preámbulos, en la Figura 10 se observa cuán fuerte es la asociación entre las variables y el tipo de clase que pertenece un hongo.

De acuerdo a la Figura 10 aquella variable que posee una mayor correlación con las clases en estudio es "odor" (0.971), siendo casi una correlación perfecta. Esto es bastante esperable, incluso sin un respaldo científico, ya que por naturaleza se puede pensar en el hecho de que el olor es un buen determinante acerca de que tan probable es que algo sea

atributo	coef.cramerV
stalk_shape	0.102
gill_attachment	0.1292
veil_color	0.1534
cap_surface	0.1969
ring_number	0.2148
cap_color	0.2184
cap_shape	0.2456
gill_spacing	0.3484
stalk_root	0.4068
habitat	0.4401
population	0.4874
bruises	0.5015
stalk_color_below_ring	0.5147
stalk_color_above_ring	0.5249
gill_size	0.54
stalk_surface_below_ring	0.5748
stalk_surface_above_ring	0.5879
ring_type	0.6033
gill_color	0.6808
spore_print	0.7526
odor	0.971

Figura 10: Tabla con los valores de los coeficientes V de Cramer para cada variable de la base de datos.

comestible o no. Además, vale la pena destacar que en otros estudios, como el que fue llevado por Sherratt, Wilkinson y Bain, el olor ha sido fuertemente asociado a la comestibilidad de los hongos, llegando a ser uno de los predictores resultante de aplicar una regresión condicional tanto para un conjunto de datos de hongos norteamericanos, como para un conjunto Europeo. En ambos casos la asociación con el olor ha sido fuerte. (Sherratt et al., 2005)

3.3. Gráficas de frecuencias con respecto a tablas de contingencias

Tal como se observó en la sub-sección anterior, cada una de las variables es determinante a la hora de clasificar un hongo, sin embargo, la importancia de cada uno de ellos se nota a simple vista. Para efectos de este estudio, se consideran los siguientes niveles de relación con respecto al valor del coeficiente V de Cramer:

En este sentido es que se procede a graficar aquellas tablas de contingencias pertenecientes a los atributos con mayor relevancia en el análisis, esto es, aquellos que posean un coeficiente mayor y bastante cercano a 0.6. En las siguientes Figuras (11, 12, 13, 14, 15, 16), se muestran gráficos de barras en los cuales se observan el número de ocurrencias de cada atributo con respecto a la clase que pertenece.

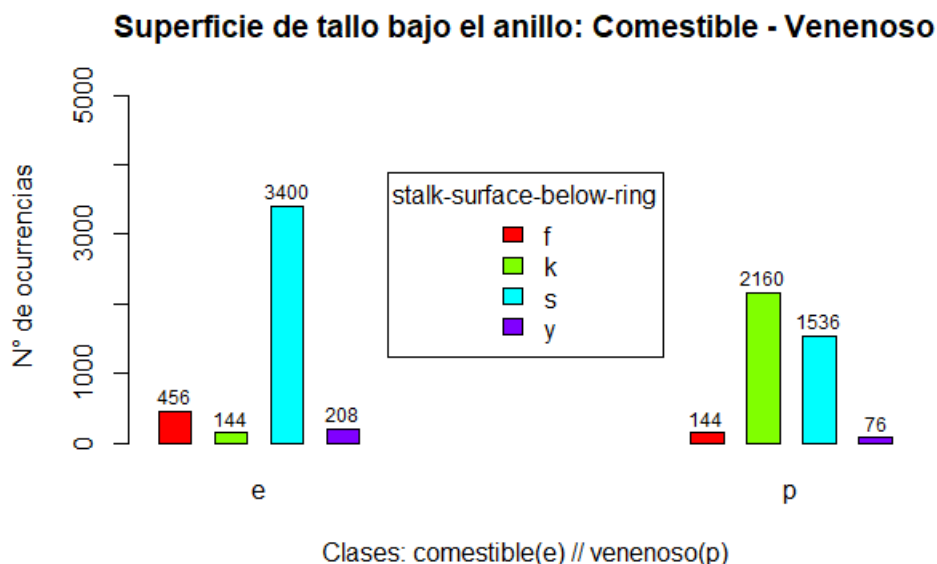


Figura 11: Gráfico de barras en base a la tabla de contingencia del atributo stalk-surface-below-ring.

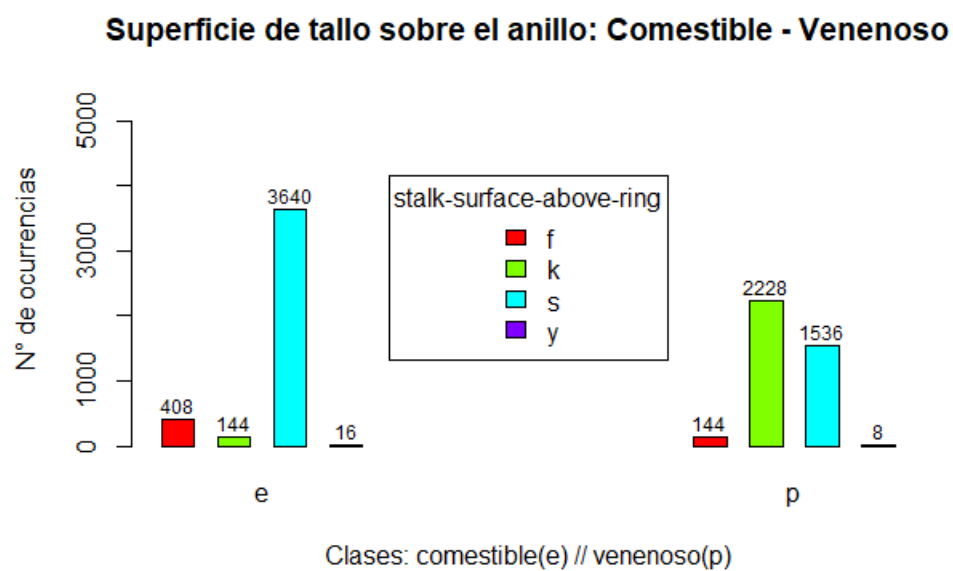


Figura 12: Gráfico de barras en base a la tabla de contingencia del atributo stalk-surface-above-ring.

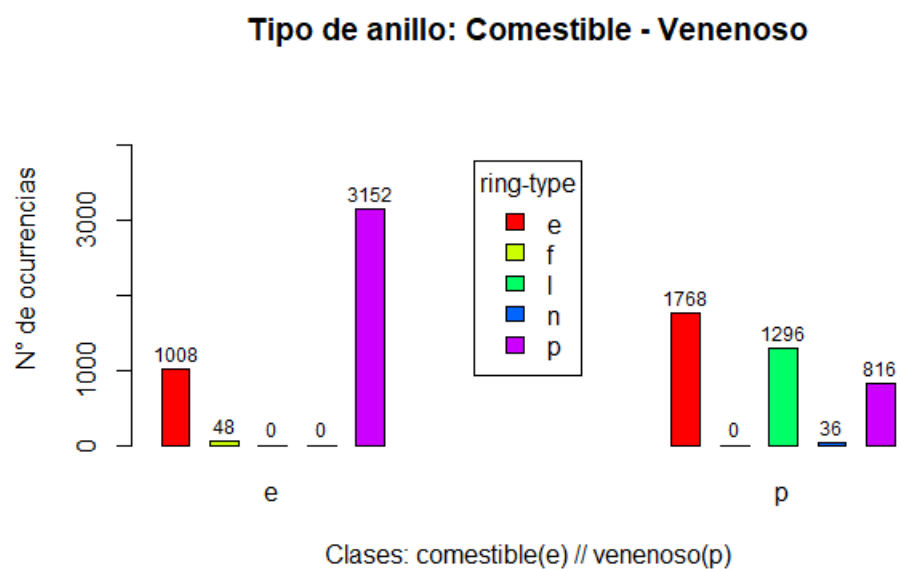


Figura 13: Gráfico de barras en base a la tabla de contingencia del atributo ring-type.

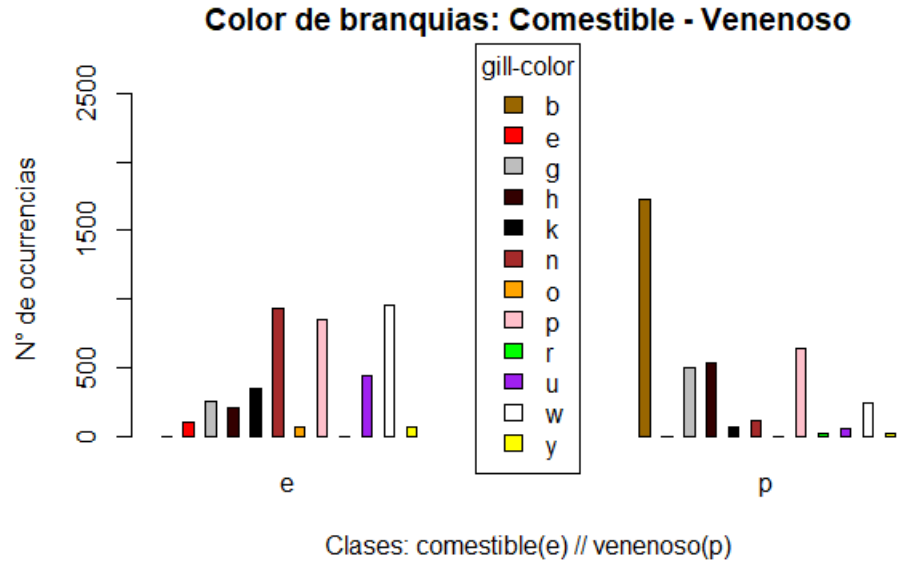


Figura 14: Gráfico de barras en base a la tabla de contingencia del atributo gill-color.

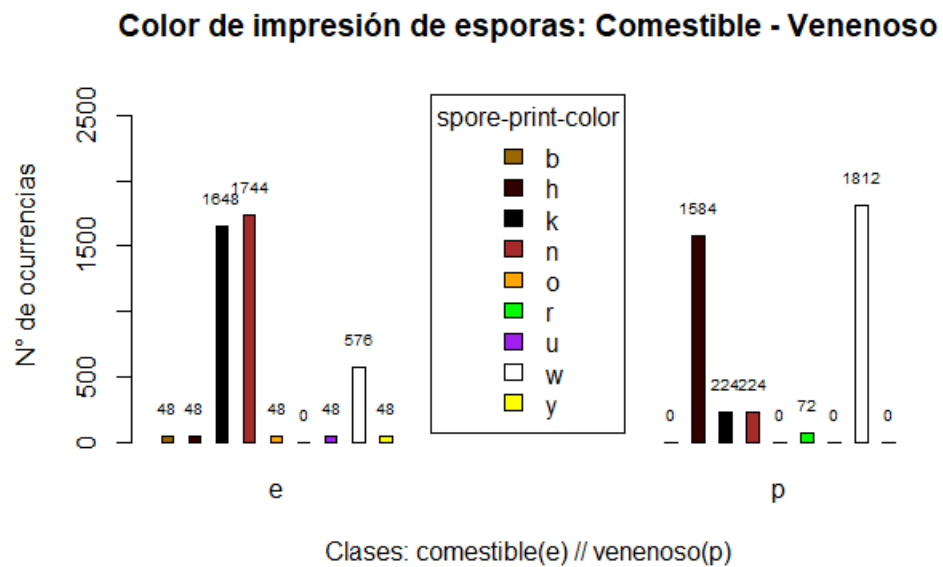


Figura 15: Gráfico de barras en base a la tabla de contingencia del atributo spore-print-color.

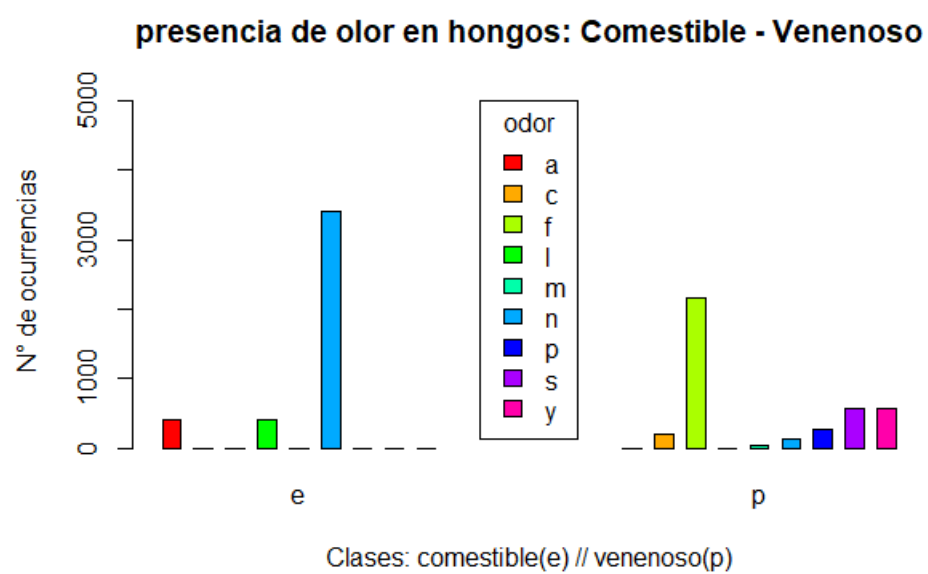


Figura 16: Gráfico de barras en base a la tabla de contingencia del atributo odor.

De la figura 11 es posible ver una tendencia de que por debajo del anillo la superficie del tipo "k" (silky) sea mayormente representativa de los hongos no comestibles. Sin embargo, esto no implica que no existan hongos comestibles que posean el mismo tipo de textura. Un caso similar ocurre con la textura "s" (smooth) que posee una tendencia marcada para el tipo de hongo comestible, aunque su cantidad en el otro grupo no debe ser desestimada, ya que de igual forma el número de aciertos encontrados se le podría considerar como alto (1536). En lo que respecta a los valores "f" (fibrous) e "y" (scally), para ambas clases las ocurrencias obtenidas son relativamente reducidas.

Para el caso de la Figura 12 es posible observar que existen tendencias similares entre los resultados de las variables stalk-surface-below-ring y stalk-surface-above-ring, pero con sutiles diferencias. Esto no sería de extrañar, ya que ambos atributos representan a diferentes partes de la misma porción del hongo, en este caso el tallo, junto con el hecho de que ambas variables poseen coeficientes V de Cramer bastante similares (0.5748 y 0.5879). Por otra parte, las diferencias que se pueden observar entre los resultados de ambas tablas de contingencia son, por ejemplo, el hecho de que hay una menor cantidad de texturas del tipo "y" (scaly) en ambos casos, a la vez que la textura "s" (smooth) es mayor en el caso de los hongos comestibles. Es por esto que se hace posible inferir, a simple vista, que la presencia de la textura scaly (escamosa) puede ser vista en mayor frecuencia en zonas del tallo inferiores al anillo. Además, tomando en consideración que el coeficiente V de Cramer de la superficie del tallo sobre el anillo es mayor que el de la inferior, y el hecho de que la cantidad de hongos comestibles con textura smooth (suave) aumentara significativamente (200 casos), se puede pensar que esta característica específica podría ser representativa de este tipo de hongos, aunque sería necesario llevar a cabo un análisis más exhaustivo.

De acuerdo al gráfico de la figura 13 y al valor del coeficiente V de Cramer obtenido por la variable ring-type es posible entrever no solo que la correlación entre esta y las clases es bastante marcada, sino que también se podría llegar a ver una relación de causalidad si es llevado a cabo un estudio que pueda demostrarlo. En específico, en primer lugar el gráfico permite observar que para los atributos "f" (flaring) y "n" (none) de la variable en estudio, en ambas clases el valor es reducido o inclusive nulo. Si bien se podría inferir que, por ejemplo, la presencia de un anillo flaring (resplandeciente) es representativo de los hongos comestibles, el hecho de que la cantidad de instancias que cumplan con esto sean cercanas a 0 hace ver que probablemente este no sea un atributo correcto para llegar a una conclusión (el mismo caso ocurre con none (ninguno) y los hongos venenosos). Sin embargo, a diferencia de lo descrito previamente, para el caso del atributo "l" (large) se puede deducir que este podría ser un

atributo correcto con el cual sea posible determinar si un hongo posee veneno, debido a que la cantidad de casos que cumplen este criterio es relativamente alto (1296), mientras que para el caso de la clase comestible no se encuentran coincidencias. Ahora bien, para el resto de los atributos solo se destaca la existencia de una gran cantidad de hongos comestibles que poseen un anillo de tipo p (pendant), pero aún así esta característica no pareciese ser exclusiva de la clase debido a que para los venenosos la cantidad de coincidencias es de igual forma alto (816 casos).

Científicamente, el color de las branquias o laminillas son un fuerte determinante en la clasificación de los hongos. El coeficiente V de Cramer para este atributo por medio de la Tabla 10 tiene un valor de 0.6808, el cual entra en un nivel de correlación fuerte. Tal como se puede observar en la Figura 14 por parte de la clase venenosa, el color más predominante de branquias es el ante (color parecido a un café claro) con más de 1500 ejemplares, el cual es una característica bastante importante para este estudio, pues en la clase comestible no existen hongos con dicho color de branquias. También es posible distinguir que ambas clases tienen un número de ejemplares bastantes parecidos que poseen un color de branquias color rosado. En general, analizando los gráficos de barras, se puede notar a simple vista que los hongos comestibles suelen ser más coloridos en lo que respecta a sus características, sin embargo no se debe afirmar de forma estricta que el hecho de ser colorido indique que el seta (parte fructífera del hongo) pueda ser consumida.

El color de impresión de esporas corresponde al depósito de polvo que cae desde el cuerpo fructífero sobre una superficie. Tal como se describió en la Figura 7 y se observó en la tabla de la Figura 10 este atributo es el segundo con mayor relevancia en lo que respecta a la determinación de la clase. Según la Tabla 10 el coeficiente de "spore-print" posee un valor de 0.7526 un número bastante alto y que por lo tanto entraría en el nivel de correlación fuerte. Ahora bien, al momento de contemplar la Figura 15 se puede distinguir que el color de impresión de esporas blanca y color chocolate son más habituales en hongos venenosos que en comestibles. No obstante, aproximadamente 575 ejemplares pertenecientes a la clase comestibles también arroja esporas de color blanco, por lo tanto no se puede concluir firmemente que solo los hongos venenosos expulsar esporas de dicho color. De forma análoga, para la clase de hongos comestibles, la impresión de esporas que más suelen expulsarse son de color negro (1648 ejemplares) y color café (1744 ejemplares). A diferencia de la clase venenosos, los hongos comestibles tienen una mayor variación de colores al momento de arrojar sus esporas, expulsando colores como amarillo, púrpura, anaranjado, negro y café.

Como se pudo observar de la Figura 16, en la gran mayoría de los hongos comes-

tibles, más de 3000 ejemplares (barra color celeste), el olor es inexistente, a excepción de algunos casos (barra roja y verde), menos de 1000 hongos, que poseen olores comparados a la almendra o el anís. Por otro lado, los hongos venenosos comúnmente presentan un olor fétido (barra color mostaza) y en algunos casos suelen tener con mayor probabilidad olores fuertes relacionados al pescado o algo picante (amoniaco). De la misma forma, se puede contemplar que la barra celeste en los hongos venenosos equivale a 120 ejemplares que no poseen olor, esto es algo que se debe tener en consideración, pues podría provocar intoxicaciones en personas que creen que todos los hongos venenosos presentan mal olor. Es por ello que a pesar de que un hongo no presente olor, existe una pequeña probabilidad que este pertenezca a la clase venenoso.

Ahora se hará un cambio de perspectiva y se analizará aquella variable que presenta un menor coeficiente de correlación. Principalmente solo se estudiará el atributo "stalk-shape" dado que los demás atributos son parecidos a este (coeficiente de correlación débil) y siguen la misma lógica de análisis. Para la familia de los hongos comestible existe un total de 4208 ejemplares, de los cuales 38 % corresponde a hongos con un tallo dilatado y 62 % a hongos con tallo estrecho. De forma análoga, para la familia de los hongos venenosos, existe un total de 3916 ejemplares de los cuales 48.5 % tiene tallo dilatado y el otro 51.8 % posee tallo estrecho. Posterior a este pequeño pero preciso análisis se puede observar que en efecto este atributo no brinda información relevante para discernir entre un hongo comestible o venenoso, y esto porque en ambas familias se tiene aproximadamente la misma cantidad de ocurrencias (expresados en %) y por lo tanto no es posible concluir de manera exacta que aquellos hongos comestibles casi siempre tendrán un tallo estrecho y los hongos venenosos un tallo dilatado (y viceversa).

Desde una mirada inferencial se encontró una serie de reglas basadas en un algoritmo denominado "MLP2LN constructivo" que permiten identificar a un hongo venenoso. A medida que las reglas se van tomando en conjunto, el nivel de precisión se acerca al 100 %. (Duch, sf). A través de la siguiente imagen (Figura 17) se puede observar como la precisión se va incrementando a medida que se van tomando en cuenta una mayor cantidad de reglas.

\mathcal{R}_1)	<code>odor=¬(almond∨anise∨none)</code>	98.52%
\mathcal{R}_2)	<code>spore-print-color=green</code>	99.41%
\mathcal{R}_3)	<code>odor=none∧stalk-surface-below-ring=scaly∧(stalk-color-above-ring=¬brown)</code>	99.90%
\mathcal{R}_4)	<code>habitat=leaves∧cap-color=white</code>	100%

Figura 17: Reglas para identificar un hongo venenoso.

4. Conclusión

El problema de determinar la comestibilidad de los hongos termina siendo tanto importante como interesante, debido a que, para dar obtener resultados se tiene que pensar, analizar y utilizar herramientas que permitan dilucidar una solución. Tal fue el caso de lo realizado en la sección de análisis del presente informe, en donde se plasmaron los datos encontrados y la información que estos implicaron. En concreto, los datos permitieron concluir que ciertas variables y atributos no resultaban provechosos para el análisis puesto que, o correspondían a una característica general de los hongos en estudio, como en el caso de la variable "veil type" y el atributo "parcial" con el cual se refleja la idea de que las observaciones de la muestra examinada tienen la característica de ser maduras, dejando de lado el tipo de velo universal (atributo "u") que denota inmadurez; o no eran representativos de las diferentes clases, como los atributos "c" (cobwebby), "s" (sheathing) y "z" (zone) de la variable "ring-type", los cuales terminaron siendo descartados. Igualmente, luego de aplicar pruebas de independencia a las variables se llegó a que todas estas estaban relacionadas con las clases de una u otra forma. Por otro lado, luego de aplicar el cálculo de la V de Cramer a las variables en estudio fue posible determinar aquellas que poseyeran una correlación más fuerte con las clases. Luego fueron comparadas cada una de estas variables en términos de cada clase en particular (tabla de contingencia), llegando a ciertas características con las que probablemente se puedan explicar los niveles de correlación, como lo es la presencia de ciertos atributos para un cierto tipo de hongo específico, o una alta cantidad de observaciones en comparación con la otra clase. Los resultados anteriores, en específico los de correlación de variables, llegan a tener sentido con respecto al problema estudiado. Esto se explica en, por un lado, los diferentes estudios con los cuales se ha llegado a soluciones similares. Tal es el caso del estudio llevado a cabo por Adamczak, Grabczewski y Zal, en el cual las reglas ahí presentadas para clasificar un hongo involucran variables como "odor", "spore-print-color", "stalk-surface-below-ring", "stalk-surface-above-ring" (Duch, sf), a las cuales se les asoció un valor fuerte dentro del cálculo de la V de Cramer. Junto al anterior, también se encuentra el estudio realizado por Sherratt, Wilkinson y Bain mencionado en la sección de análisis, en el cual también se remarcó la importancia del olor para discernir si un hongo es comestible o no (Sherratt et al., 2005)

En base a los objetivos planteados en un inicio, efectivamente se logró entender de manera eficaz cada uno de los atributos pertenecientes a las familias de los hongos, se

llevo a cabo una tabla en donde se describe de manera precisa que significa cada variable y en ocasiones su importancia en el estudio de clasificación de hongos. También fue necesario conocer todos aquellos atributos que puede tomar una variable, ya sea los distintos colores, olores, texturas, formas, entre otras. Con esta base de conocimiento fue posible dilucidar de manera correcta la anatomía y comportamientos que posee tanto la familia de hongos comestibles como venenosos. Posterior a esta investigación exhaustiva de las características de los hongos, se pudo llevar a cabo un adecuado análisis descriptivo acerca de los hongos y cuan determinantes eran sus variables. Por parte del estudio inferencial se recopiló información de un paper (Duch, sf) acerca de la identificación de un hongo comestible, en donde surgen 4 reglas para su determinación. Y como último objetivo, se logró aplicar una interpretación a los resultados de forma correcta por medio de distintos métodos ("prueba de independencia de Chi-cuadrado" y "Método V de Cramer") que permitieron que los resultados obtenidos concordaban efectivamente con la información que los gráficos de frecuencias arrojaban.

Finalmente, con el fin de dar cuenta del trabajo realizado, se puede resumir la experiencia de análisis de los datos en una serie de actividades concretas, las cuales son: generar las tablas de contingencias, analizarlas en búsqueda de variables sin aportes significativos al estudio del problema, realizar pruebas de independencia chi-cuadrado y Fisher de acuerdo a lo necesario para cada caso, determinar correlación por medio de la V de Cramer, análisis de las características relevantes de atributos dentro de aquellas variables con mayor y menor grado de correlación, y por último, a través del estudio perteneciente al paper (Duch, sf), se llegó a un algoritmo basado en redes neuronales con el cual fue posible, por un lado, comprobar la importancia de ciertas variables como el olor, y por el otro, la definición de ciertas reglas con las cuales se puede resolver el problema.

Bibliografía

- Duch, W. (s.f.). A hybrid method for extraction of logical rules from data. Recuperado 22 de abril de 2022 en <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.32.5136&rep=rep1&type=pdf>.
- Fondation Miron-Royer Inc (2018). Identification criteria.
- Furci, G. and Magnasco, C. (2017a). Fungi austral - guía de campo de los hongos más vistosos de chile. (p. 20). Recuperado 16 de abril de 2022 en https://issuu.com/hongos.cl/docs/lib_hongos_baja?viewMode=magazine.
- Furci, G. and Magnasco, C. (2017b). Fungi austral - guía de campo de los hongos más vistosos de chile. (p. 21). Recuperado 16 de abril de 2022 en https://issuu.com/hongos.cl/docs/lib_hongos_baja?viewMode=magazine.
- Jara, J. L. and Köhler, J. (2021a). Inferencia y modelos estadísticos. Universidad de Santiago de Chile. (p. 114). Recuperado 19 de abril de 2022.
- Jara, J. L. and Köhler, J. (2021b). Inferencia y modelos estadísticos. Universidad de Santiago de Chile. (p. 115). Recuperado 19 de abril de 2022.
- Magaña, S. (2019). Por qué las setas son indispensables para conservar el planeta. National Geographic España. [Online] Recuperado el 15 de abril de 2022 en https://www.nationalgeographic.com.es/ciencia/por-que-setas-son-indispensables-para-conservar-planeta_14912.
- Osorio, U. R. (2021). Estructura de los hongos. ecologiaverde.com. Recuperado 16 de abril de 2022 en <https://www.ecologiaverde.com/estructura-de-los-hongos-3676.html>.
- Sherratt, T. N., Wilkinson, D. M., and Bain, R. S. (2005). Explaining dioscorides' "double difference": why are some mushrooms poisonous, and do they signal their unprofitability?. *The American Naturalist*. 166(6), p. 770. Recuperado 20 de abril de 2022 en http://www.davidmoore.org.uk/21st_Century_Guidebook_to_Fungi_PLATINUM/REPRINT_collection/Sherratt_etal_ecology_mushroom_poisons.pdf.
- Torres, A. (2013). Asociación hongo y planta beneficia a cultivos. [Online] Recuperado el 16 de abril de 2022 en <https://www.ucr.ac.cr/noticias/2013/07/26/asociacion-hongo-y-planta-beneficia-a-cultivos.html>.

Tullii, P. A. . M. (2021). Fungi. Monaco Nature Encyclopedia. Recuperado 17 de abril de 2022 en <https://www.monaconatureencyclopedia.com/fungi/?lang=es>.

UCI (s.f.). Uci machine learning repository: Mushroom data set. Machine Learning Repository. Recuperado 13 de abril de 2022 en <https://archive.ics.uci.edu/ml/datasets/mushroom>.