



Instituto Tecnológico Autónomo de México
Departamento de Administración
DECISIONES DE NEGOCIOS BASADAS EN DATOS

¿Cómo predecir si un hongo es venenoso dependiendo de sus características físicas y así prevenir las muertes causadas por intoxicación?

Midterm Project

Integrantes

Ana Paula Graf Phillips
María Fernanda Rodríguez González 173176
Gabriela Santibañez 174099
José Ignacio Moreno Novaro 165183
Eduardo Pesqueira Zorrilla 176065

Introducción del tema y contexto.

Las intoxicaciones por el consumo de hongos silvestres son un problema global con una tasa de mortalidad alta, pues algunos estudios indican que aproximadamente el 15% de las intoxicaciones por hongos silvestres tienen consecuencias fatales. La intoxicación se debe a la ingestión de las sustancias tóxicas presentes en algunos hongos que tienen efectos nocivos para la salud. Los síntomas pueden variar desde una leve molestia gastrointestinal hasta la muerte; sin embargo, sorprendentemente, la razón más común que causa estas intoxicaciones se debe a la identificación errónea por la gran semejanza en términos de color y morfología general que tienen especies de hongos tóxicos con las especies de hongos comestibles.

Es importante tener en cuenta que los hongos que crecen en la tierra son más peligrosos que los que crecen en árboles vivos y los hongos en la tierra de los bosques son más peligrosos que los hongos en los patios. Además de que las toxinas de los hongos no se eliminarán al cocinarlos. Los síntomas más comunes de una intoxicación grave son los del síndrome faloide, que se refiere a dolor abdominal, vómitos y diarreas, además de ictericia, sangrado y finalmente coma hepático. Los síntomas pueden aparecer inmediatamente o incluso tardar unas horas, pero generalmente mientras más rápido empiecen los síntomas la intoxicación es menos grave.

En México, por ejemplo, entre 2005 y 2013 hubo, solo en las comunidades indígenas de Los Altos de Chiapas, 85 casos con un saldo de 31 muertos. Si bien esta no es una de las principales causas de mortalidad en Chiapas, es un problema de salud que afecta una práctica cultural ancestral, como lo es la recolección de hongos silvestres, y afecta la economía local.

Para poder prevenir el envenenamiento por hongos, es necesario familiarizarse con la apariencia y características de los hongos comestibles que se buscan recolectar y sus diferencias con cualquier especie de hongos tóxicos que tengan una apariencia similar. Adicionalmente, la seguridad de comer hongos silvestres también puede depender de los métodos de preparación antes de ingerirlos.

Con base en esto, utilizamos una base de datos que contiene descripciones de las características físicas de muestras correspondientes a 23 especies de hongos extraídas de la Guía de campo de la Sociedad Audubon sobre hongos norteamericanos en 1981, con la finalidad de encontrar algún tipo de relación entre las características de un hongo y su toxicidad. En la base, cada especie se identifica como: definitivamente comestible, definitivamente venenosa o de toxicidad desconocida.

El propósito de este estudio es ayudar a cuantificar el riesgo que existe al recolectar un hongo debido a sus características físicas para su consumo y se prevé que los resultados podrán ayudar a las comunidades cuyas economías se benefician del consumo de hongos a reducir su número de accidentes anuales.

Pregunta a resolver: ¿Cómo determinar si un hongo es venenoso dependiendo de sus características y así prevenir las muertes causadas por intoxicación?

Importancia de la pregunta

Solamente en el estado de Chiapas, existen aproximadamente 13 mil especies de hongos vegetales. De todos estos, solamente se conocen 0.5% de ellos, de los cuales 40 son comestibles y los otros 6 son venenosos. Esta industria de la venta de hongos no ha crecido como podría, ya que solo 54% de los mexicanos consume

hongos en el año. De esas 13 especies de hongos seguro que hay muchos más que podrían ser populares pero no se ha llegado a ellos porque no sabemos si son o no venenosos.

Con la pregunta que queremos resolver dados ciertos atributos de los hongos, se va a poder identificar con mucha precisión si un hongo es comestible, venenoso o no recomendado, y así evitar intoxicaciones y muertes a la hora de probar hongos y así poder crecer esta industria.

Descripción de la base de datos:

Recopilación de data

La base de datos que se ha obtenido para este proyecto fue extraída de *University of California Irvine - Machine Learning Repository*. Este conjunto de datos es llamado “Mushroom Dataset” e incluye descripciones de muestras correspondientes a 23 especies de hongos con branquias de la familia *Agaricus* y *Lepiota*. Cada especie se identifica como definitivamente comestible, definitivamente venenosa o de comestibilidad desconocida y no se recomienda. Esta última clase se combinó con la venenosa.

Así mismo, este registro de hongos es extraído de *The Audubon Society Field Guide to North American Mushrooms* (1981). G. H. Lincoff (Pres.), Nueva York: Alfred A. Knopf

Se cuenta con 22 variables dentro de la base de datos, las cuales, cada una aporta información acerca de las características físicas propias de un hongo. Ver *Anexo 1* donde se desglosa una breve descripción de cada variable.

¿Porqué la base de datos es adecuada para resolver la pregunta de investigación?

Para el análisis y solución de la pregunta se ha decidido trabajar con una base de datos con información cualitativa. Se cree que la base de datos elegida contiene las características necesarias para poder predecir la naturaleza de un hongo dadas sus propiedades físicas. Trabajar con una base de datos que contiene 23 variables y 126 atributos cualitativos es suficiente para poder convertir nuestra investigación en una confiable. La pregunta a resolver tiene el objetivo de, dadas ciertas características físicas de un hongo, predecir si un hongo es venenoso o no. Los datos recopilados son adecuados ya que cualquier tipo de hongo podrá ser identificado por las variables que se han definido anteriormente.

Análisis preliminar de datos

El objetivo es predecir si el hongo es comestible o no, para llevar a cabo esto sabemos que es un trabajo de aprendizaje supervisado, específicamente una clasificación. Primeramente, se debía de buscar aquellas variables que aportaran mayor información en el proceso de clasificación, por lo cual se modeló un árbol de clasificación. Después de haber dividido nuestro dataset entre un set de entrenamiento y pruebas, se buscó el árbol óptimo, es decir, aquel con menor entropía en las hojas. En el **anexo 7** se muestran los hiper-parámetros del árbol óptimo y se muestra el árbol en el **anexo 8**.

Las variables que aportaron más información fueron las que se presentan en la siguiente tabla:

	importance	feature
0	odor_none	61.241514641%
1	stalk_root_club	17.804251486%
2	stalk_surface_below_ring_scaly	10.46435801%
3	spore_print_color_green	3.4494247%
4	gill_spacing	2.43358111%
5	bruises	2.315494498%
6	ring_type_pendant	1.545870911%
7	stalk_root_bulbous	0.628568999%
8	cap_color_white	0.105303975%
9	gill_size	0.01163167%

El árbol tiene una raíz con la variable que indica que los hongos no tienen olor, seguido de la raíz ruta *club* y la superficie del tallo *scally*.

Las 3 primeras características aportan un 88% de información, se hicieron unas gráficas donde esta información se ve clara véase en el **anexo 9**.

El siguiente paso consistirá en realizar un modelo de Random Forest, en donde se busque la predicción de los datos y se examinarán las métricas de desempeño pertinentes.

Referencias

Consultado 30/Jun/2021:

<https://archive.ics.uci.edu/ml/datasets/mushroom>

Consultado 04/Jul/2021

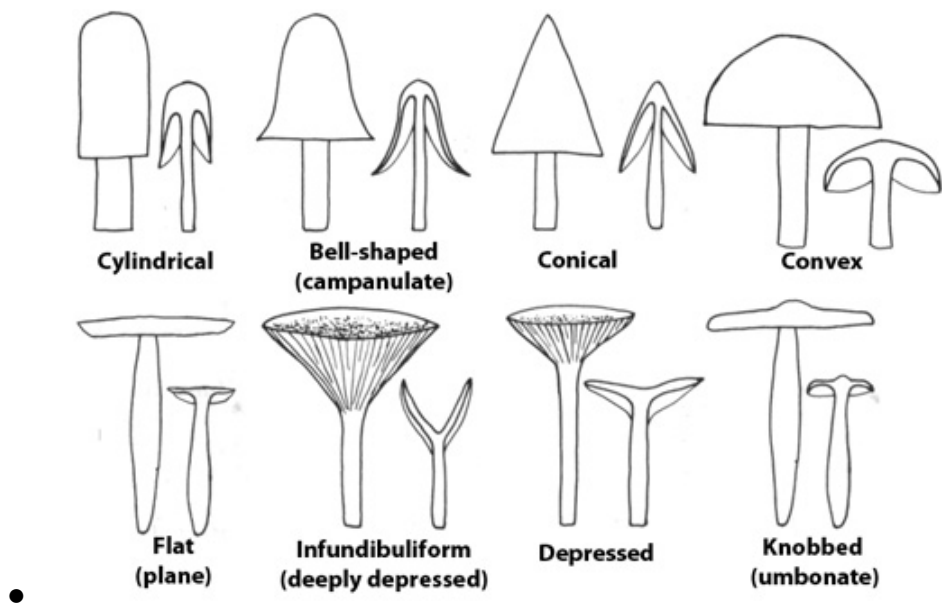
<https://noticieros.televisa.com/ultimas-noticias/consumo-hongos-chiapas-tradicion-ancestral/>

Consultado 05/Jul/2021:

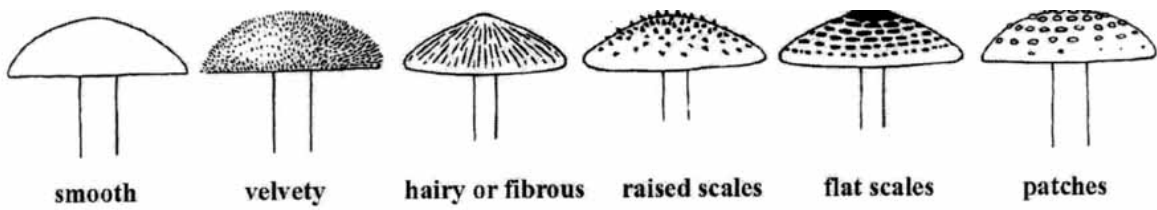
<https://es.familydoctor.org/condicion/intoxicacion-alimentaria-envenenamiento-por-hongos/>

Anexos

Anexo 1: *Forma de capuchón del hongo*



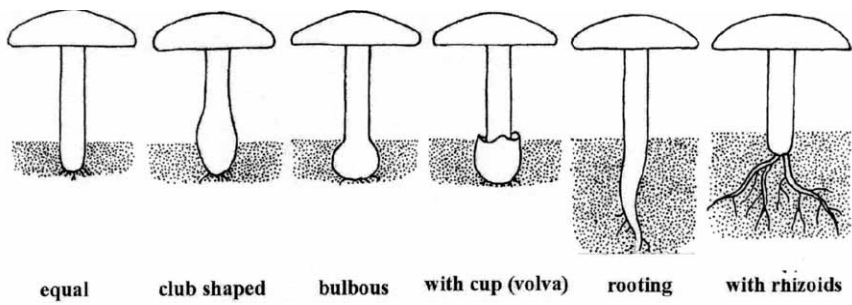
Anexo 2:



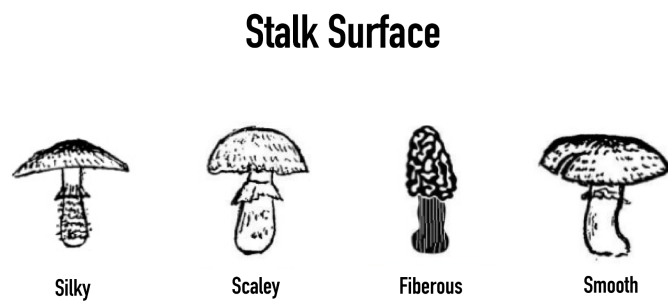
Anexo 3:



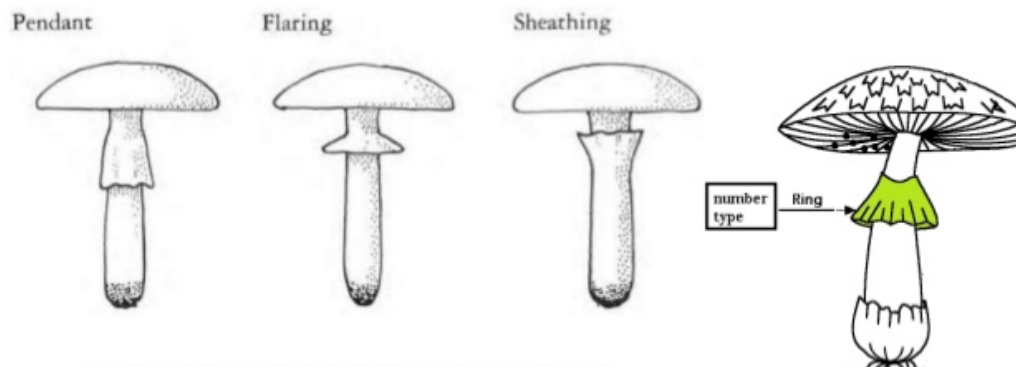
Anexo 4:



Anexo 5:



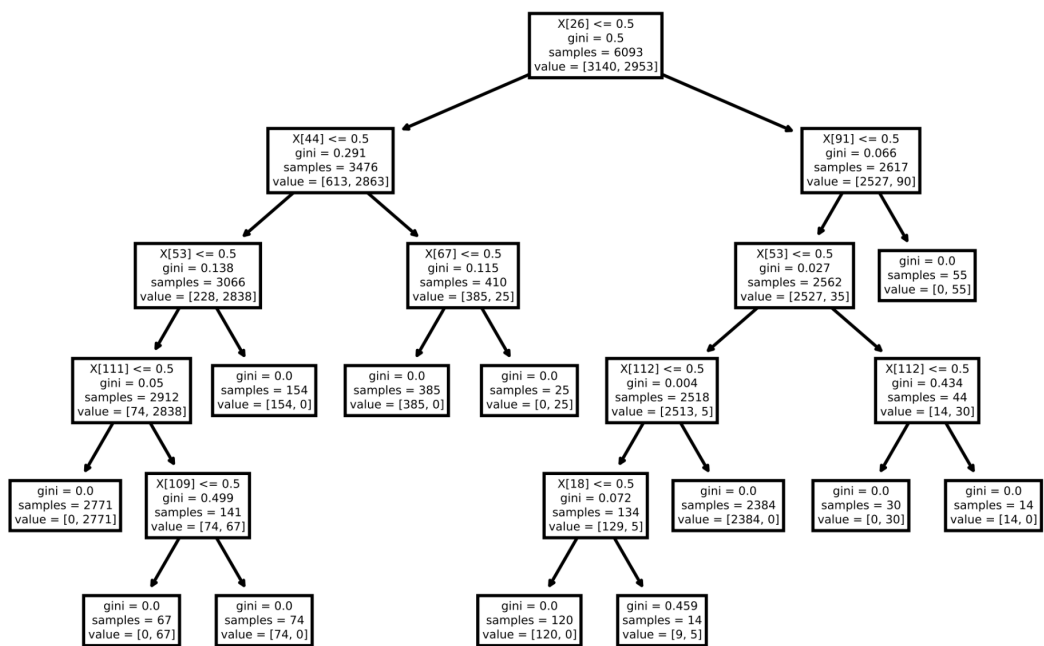
Anexo 6:



Anexo 7:

[18]: DecisionTreeClassifier(max_depth=5, min_samples_leaf=7, random_state=5432)

Anexo 8:

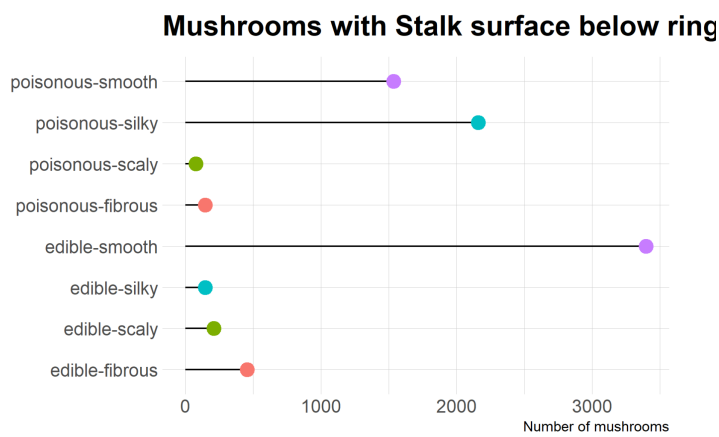
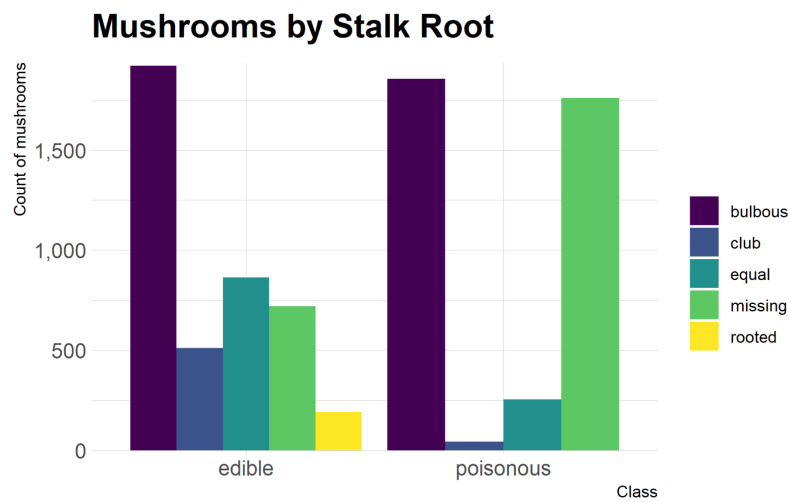
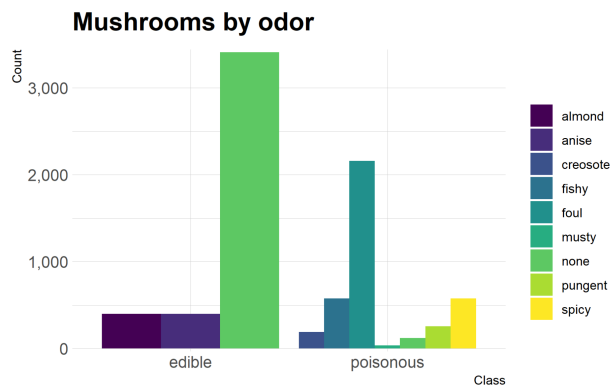


```

|--- odor_none <= 0.50
| |--- stalk_root_club <= 0.50 => 0
| | |--- stalk_surface_below_ring_scaly <= 0.50
| | | |--- gill_spacing <= 0.50
| | | |--- class: 1
| | | |--- gill_spacing > 0.50
| | | |--- bruises <= 0.50
| | | | |--- class: 1
| | | |--- bruises > 0.50
| | | |--- class: 0
| | |--- stalk_surface_below_ring_scaly > 0.50
| | |--- class: 0
| |--- stalk_root_club > 0.50 = 1
| | |--- stalk_color_below_ring_cinnamon <= 0.50
| | |--- class: 0
| | |--- stalk_color_below_ring_cinnamon > 0.50
| | |--- class: 1
|---1 odor_none > 0.50
| |---2 spore_print_color_green <= 0.50
| | |---3 stalk_surface_below_ring_scaly <= 0.50
| | | |--- 4 gill_size <= 0.50
| | | | |--- 5 cap_color_white <= 0.50
| | | | |--- 6 class: 0
| | | |--- cap_color_white > 0.50
| | | | |--- class: 0
| | | |--- gill_size > 0.50
| | | |--- class: 0
| | |--- stalk_surface_below_ring_scaly > 0.50
| | | |--- gill_size <= 0.50
| | | |--- class: 1
| | | |--- gill_size > 0.50
| | | |--- class: 0
| |--- spore_print_color_green > 0.50
| |--- class: 1

```

Anexo 9:



Anexo 10

Variable	Descripción
Clase	Si es comestible o no
Forma de capuchón	Figura física de capuchón del hongo, ver <i>Anexo 1</i>
Superficie de capuchón	Textura de capuchón del hongo, ver <i>Anexo 2</i>
Color de capuchón	Café, blanco, amarillo, gris, morado, rosa, verde, canela, beige, rojo
Moretones	Manchas - Verdadero o falso, ver <i>Anexo 3</i>
Olor	Referencia a 9 tipos diferentes de olores del hongo (spicy, anís, almendra, fishy, creosota, foul, rancio, inoloro, acre)
Accesorio de branquias	Adjuntas, descendentes, libre o mellado
Separación de branquias	Distantes cercanas o amalgamadas
Tamaño de las branquias	anchas o estrechas
Color de las branquias	negro, café, beige, chocolate, gris, verde, naranja, rosa, morado, rojo, blanco, amarillo
Forma del tallo	alargandose, estrechandose.
Raíz del tallo	figura física de raíz del tallo, ver <i>Anexo 4</i>
Superficie sobre el anillo	ver <i>Anexo 5</i>
Superficie debajo del anillo	ver <i>Anexo 5</i>
Color del tallo sobre el anillo	cafe, beige, canela, gris, naranja, rosa, rojo, blanco, amarillo
Color del tallo debajo del anillo	cafe, beige, canela, gris, naranja, rosa, rojo, blanco, amarillo
Tipo de velo	parcial, universal
Color del velo	cafe, naranja, blanco o amarillo
Número de anillo	0, 1, 2,
Tipo de anillo	Ver <i>Anexo 6</i>
Color de impresión de esporas	Negro, cafe, beige, chocolate, verde, naranja, morado, blanco o amarillo
Población	abundante, numerosa, scattered, solitaria, "several", "clustered"
Hábitat	pasto, hojas, madera, deshechos, prado, rutas, urbano