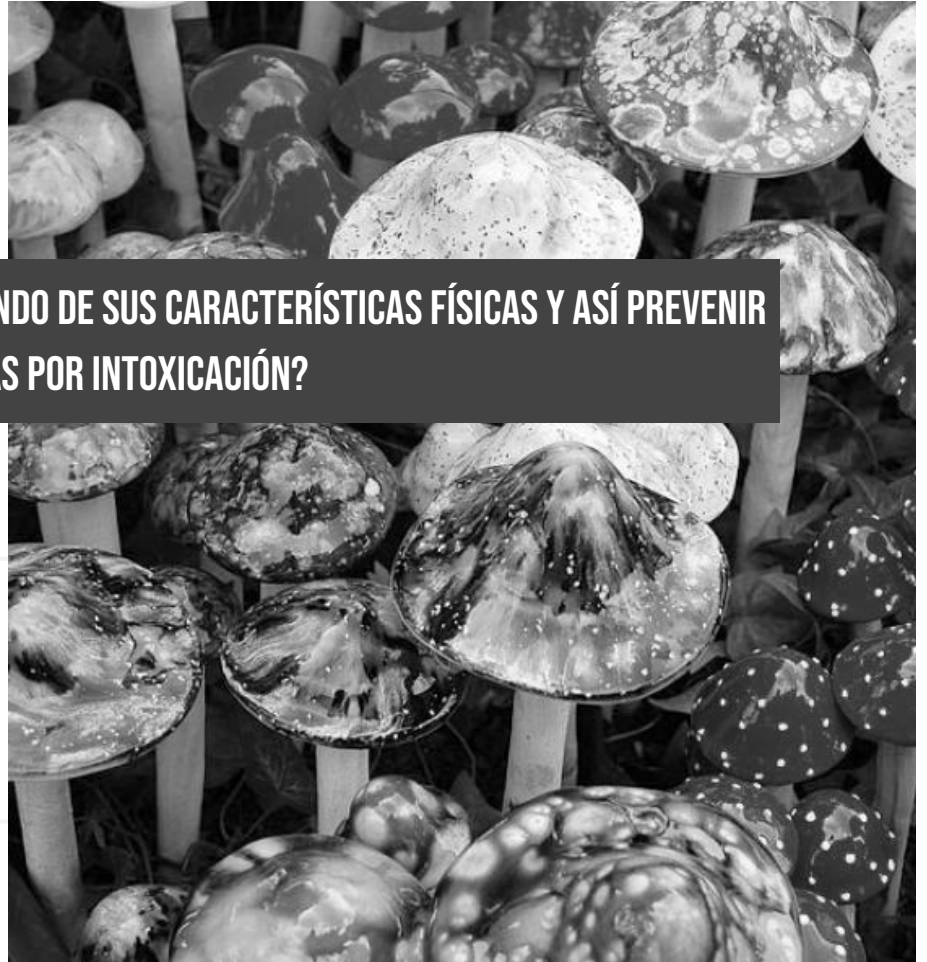


# MIDTERM PROJECT

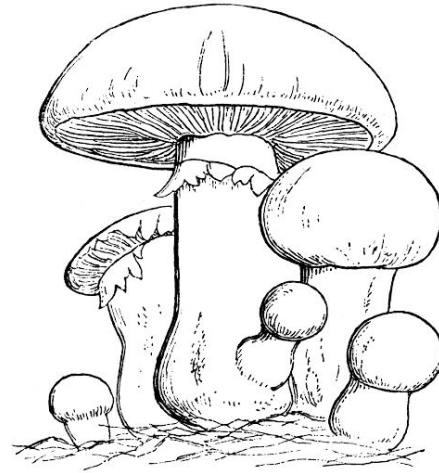
**¿CÓMO PREDECIR SI UN HONGO ES VENENOSO DEPENDIENDO DE SUS CARACTERÍSTICAS FÍSICAS Y ASÍ PREVENIR LAS MUERTES CAUSADAS POR INTOXICACIÓN?**

Fernanda Rodríguez  
Ana Paula Graf  
Eduardo Pesqueira  
Gabriela Santibañez  
José Ignacio Moreno



# ¿INTOXICACIÓN POR HONGOS?

- ¿Qué es la intoxicación por hongos y por qué es un tema importante?
- ¿Cómo se diagnostica el envenenamiento por hongos?



# INTRODUCCIÓN

- ¿Se puede prevenir o evitar la intoxicación por hongos?
- ¿Cómo se puede ayudar a las comunidades que recolectan hongos para consumir y vender?



# IMPORTANCIA DE LA PREGUNTA

# 02

- Sólo en Chiapas 13k especies
- Conocemos 0.5% de ellos
- 54% de los mexicanos consumen hongos
- Crecer la industria
- Evitar intoxicaciones y muertes



# DESCRIPCIÓN DE NUESTRA BASE DE DATOS

- La base de datos que se ha obtenido para este proyecto fue extraída de *The University of California Irvine - Machine Learning Repository*.
- Contiene descripciones de muestras correspondientes a 23 especies de hongos con branquias de la familia Agaricus y Lepiota
- Se cuenta con 23 variables dentro de la base de datos, las cuales, cada una aporta información acerca de las características físicas propias de un hongo.

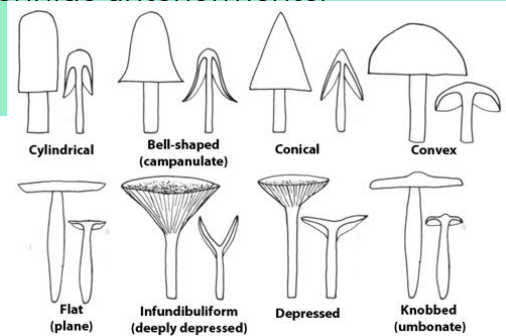
	Variable	Descripción
1	Clase	Si es comestible o no
2	Forma de capuchón	Figura física de capuchón del hongo, ver <i>Anexo 1</i>
3	Superficie de capuchón	Textura de capuchón del hongo, ver <i>Anexo 2</i>
4	Color de capuchón	Café, blanco, amarillo, gris, morado, rosa, verde, canela, beige, rojo
5	Moretones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



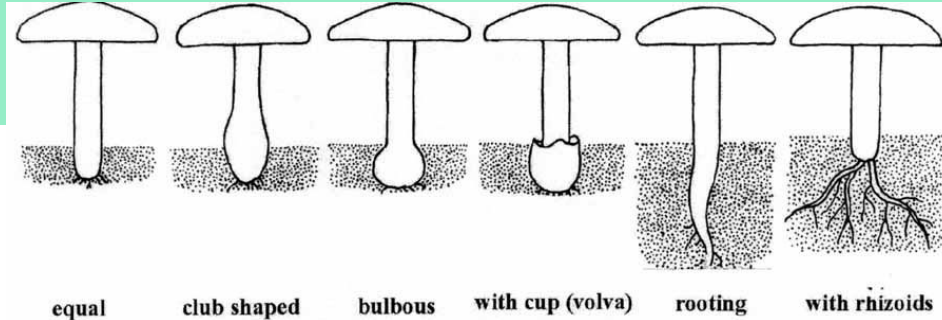
# ¿PORQUÉ LA BASE DE DATOS ES ADECUADA PARA RESOLVER LA PREGUNTA DE INVESTIGACIÓN?

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.

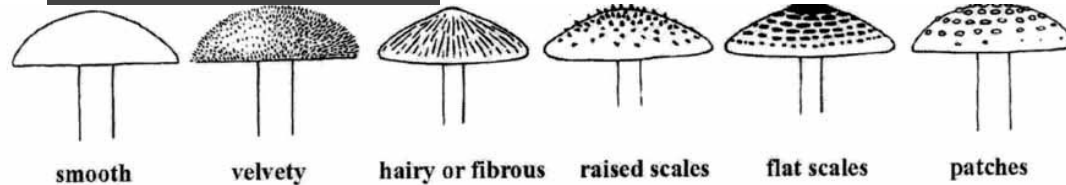
La pregunta a resolver tiene el objetivo de, dadas ciertas características físicas de un hongo, predecir si un hongo es venenoso o no. Los datos recopilados son adecuados ya que cualquier tipo de hongo podrá ser identificado por las variables que se han definido anteriormente.



1. Forma de capuchón

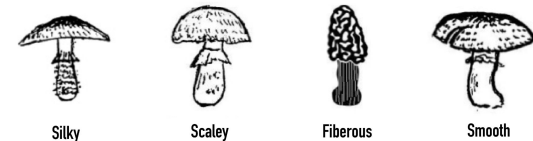


2. Forma de la raíz del tallo



3. Superficie da capuchón

4. Superficie sobre el anillo



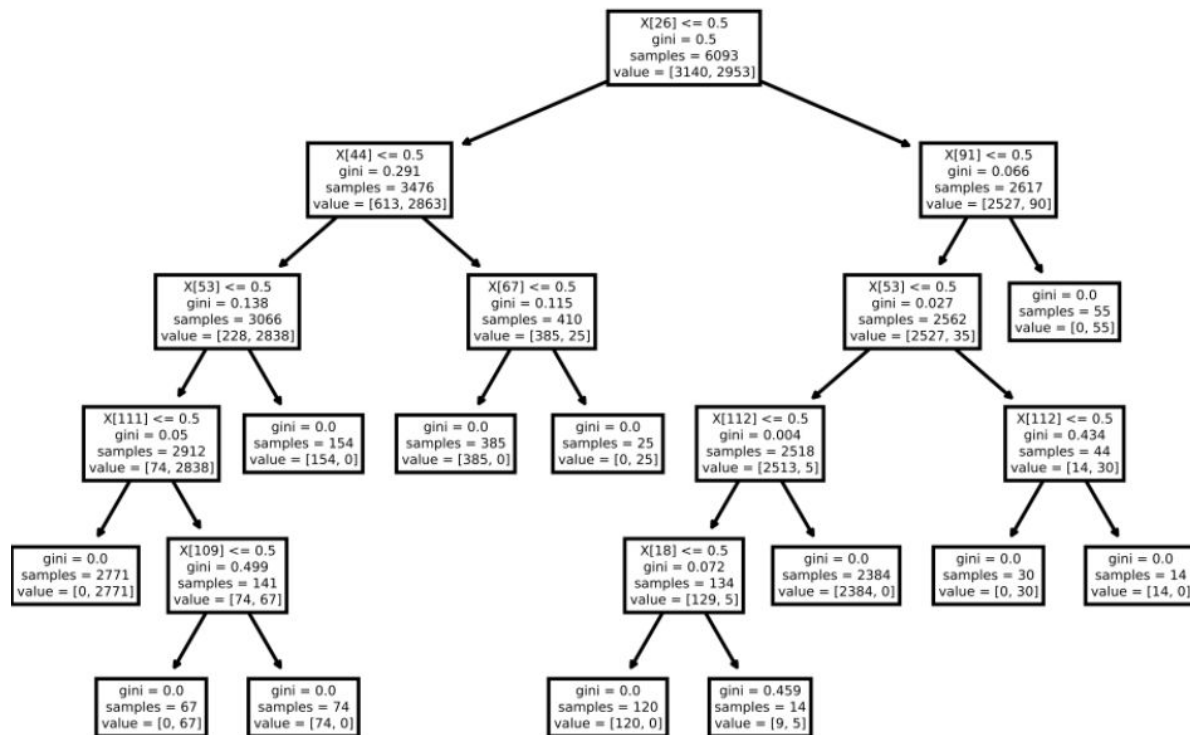
# 03

## ANÁLISIS PRELIMINAR



# DECISION TREE

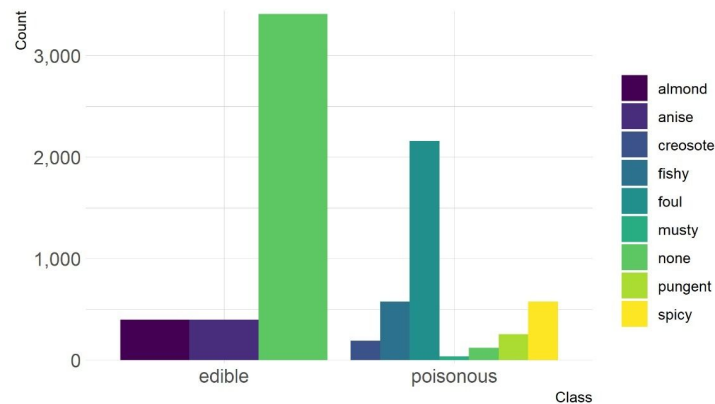
	importance	feature
0	odor_none	61.241514641%
1	stalk_root_club	17.804251486%
2	stalk_surface_below_ring_scaly	10.46435801%
3	spore_print_color_green	3.4494247%
4	gill_spacing	2.43358111%
5	bruises	2.315494498%
6	ring_type_pendant	1.545870911%
7	stalk_root_bulbous	0.628568999%
8	cap_color_white	0.105303975%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



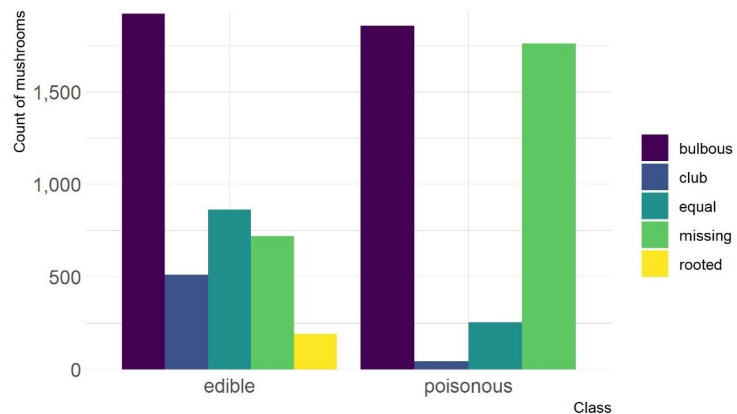


# VISUALIZACIONES

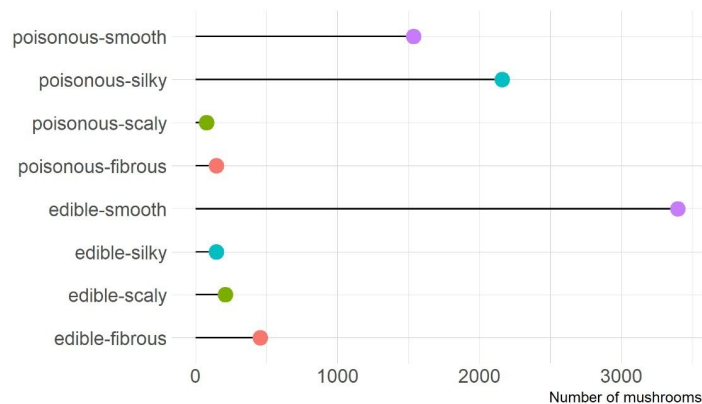
## Mushrooms by odor



## Mushrooms by Stalk Root



## Mushrooms with Stalk surface below ring



# SIGUIENTES PASOS:



## Random Forest

Se buscará hacer un modelo de Random Forest con ayuda de herramientas de python (*scikit-learn*).

Donde se buscará maximizar la cobertura, es decir, del universo posible de nuestra clase positiva, identificar la mayoría correctamente.

- $\text{recall/cobertura} = \text{TP}/(\text{TP} + \text{FP})$

Predicción/ Real	Hongo venenoso	Hongo no venenoso
Hongo venenoso	Verdaderos positivos	Falsos positivos
Hongo no venenoso	Falsos negativos	Verdaderos negativos