

# Informe de proyecto de Identificación de Setas 🍄

## Proyecto de Identificación de Setas: Transparencia y Seguridad

Este documento detalla el desarrollo de una aplicación para identificar setas, clasificándolas como comestibles o venenosas. Nuestro objetivo principal es proporcionar una herramienta útil y, sobre todo, segura para el usuario.

### 1. El Desafío Inicial: Datos "Demasiado Perfectos"

Comenzamos con un dataset de hongos conocido por la comunidad de Machine Learning. Durante las primeras pruebas, nuestros modelos alcanzaban un 100% de precisión, lo cual es inusual y a menudo indica un problema.

**La Causa: Fugas de Datos.** Identificamos que algunas características del dataset (como el olor o el color de las esporas) eran tan predictivas que el modelo no necesitaba "aprender" realmente, simplemente seguía reglas directas que correlacionaban casi perfectamente con la comestibilidad. En un escenario real, estas características pueden ser difíciles de observar con precisión por un no experto.

**Nuestra Solución:** Para crear un modelo más realista y seguro, decidimos intencionadamente **excluir** las características altamente discriminatorias y difíciles de percibir con fiabilidad (como el olor y el color de las esporas). Esto forzó al modelo a aprender de patrones más sutiles y accesibles, haciendo que el desafío de la clasificación fuera más auténtico.

### 2. Selección de Características para la Interfaz de Usuario

Pensando en la experiencia del usuario final en nuestra aplicación Streamlit, seleccionamos cuidadosamente las características que una persona puede observar y reportar fácilmente sin ser un micólogo experto.

Las características elegidas son:

- **cap-shape** (forma del sombrero)
- **cap-surface** (superficie del sombrero)
- **cap-color** (color del sombrero)
- **bruises** (si tiene moretones al tocarla)
- **gill-color** (color de las láminas)
- **stalk-shape** (forma del tallo)
- **stalk-surface-above-ring** (superficie del tallo por encima del anillo)
- **stalk-surface-below-ring** (superficie del tallo por debajo del anillo)
- **stalk-color-above-ring** (color del tallo por encima del anillo)
- **stalk-color-below-ring** (color del tallo por debajo del anillo)
- **veil-color** (color del velo)
- **ring-number** (número de anillos)
- **ring-type** (tipo de anillo)

- **population** (cómo crece: individual, disperso, etc.)
- **habitat** (dónde crece: en el bosque, caminos, etc.)

### 3. El Modelo Elegido: Árbol de Decisión

Evaluamos varios algoritmos de Machine Learning, incluyendo K-Nearest Neighbors (KNN), Bagging, AdaBoost y Gradient Boosting. Aunque la mayoría ofrecieron un rendimiento excelente incluso con las características seleccionadas, elegimos el **Árbol de Decisión** por sus ventajas clave:

- **Transparencia y Explicabilidad:** El Árbol de Decisión funciona como una serie de preguntas simples que llevan a una decisión. Podemos ver y entender el "razonamiento" del modelo paso a paso.
- **Confianza del Usuario:** Al ser transparente, podemos explicar al usuario por qué el modelo llegó a una conclusión específica (ej. "Tu seta es venenosa porque no tiene moretones y su lámina es de color X"). Esto genera confianza, vital cuando se trata de seguridad.
- **Alto Rendimiento y Seguridad:** El modelo de Árbol de Decisión optimizado demostró una precisión muy alta en la clasificación de setas comestibles y, crucialmente, minimizó los falsos negativos (clasificar una seta venenosa como comestible), que es el error más peligroso.

### 4. Nuestro Proceso de Desarrollo (Resumen Técnico)

1. **Carga y Preprocesamiento:** Importamos los datos, identificamos y eliminamos columnas irrelevantes o con fugas de datos.
2. **Codificación de Características:** Transformamos las características categóricas seleccionadas en un formato numérico usando **One-Hot Encoding** para que el modelo pudiera procesarlas.
3. **División de Datos:** Separamos el dataset en conjuntos de entrenamiento (80%) y prueba (20%) para evaluar el modelo de forma imparcial.
4. **Optimización de Hiperparámetros:** Utilizamos **GridSearchCV** para encontrar la mejor configuración (hiperparámetros) para cada tipo de modelo, asegurando su máximo rendimiento.
5. **Evaluación Rigurosa:** Analizamos métricas clave como la precisión, el recall y el F1-score en los conjuntos de entrenamiento y prueba para entender el rendimiento de cada modelo.
6. **Guardado del Modelo:** Guardamos el Árbol de Decisión entrenado y el codificador de etiquetas para su uso posterior en la aplicación Streamlit.

Este enfoque nos ha permitido desarrollar una herramienta de identificación de setas precisa, segura y transparente, diseñada pensando en el usuario final.

---

# Guía de Características de Setas para Identificación

Para ayudarte a usar nuestra aplicación de identificación de setas, aquí te explicamos las características clave que te pediremos observar en la seta. Tu precisión al describir estas características es fundamental para obtener una predicción fiable.

---

## Características del Sombrero (Cap)

- **Forma del Sombrero (cap-shape):**
    - Campaniforme (forma de campana)
    - Cónica (forma de cono)
    - Convexa (redondeada, curva hacia afuera)
    - Plana (completamente plana)
    - Umbonada (con un bulto en el centro)
    - Hundida (con una depresión en el centro)
  - **Superficie del Sombrero (cap-surface):**
    - Fibrosa (con fibras visibles)
    - Acanalada (con surcos o canales)
    - Escamosa (con escamas o parches)
    - Lisa (sin texturas pronunciadas)
  - **Color del Sombrero (cap-color):**
    - Marrón
    - Beige
    - Canela
    - Gris
    - Verde
    - Rosa
    - Morado
    - Rojo
    - Blanco
    - Amarillo
- 

## Características de las Láminas (Gills)

- **Color de las Láminas (gill-color):**
  - Negro
  - Marrón
  - Beige
  - Chocolate
  - Gris
  - Verde

- Naranja
  - Rosa
  - Morado
  - Rojo
  - Canela
  - Blanco
  - Amarillo
- 

## **Características del Tallo (Stalk)**

- **Moretones (bruises):**
  - Sí (se magulla o cambia de color al manipularla)
  - No (no se magulla)
- **Forma del Tallo (stalk-shape):**
  - Ensanchado (se ensancha hacia la base)
  - Ahusado (se estrecha hacia la base)
- **Superficie del Tallo por Encima del Anillo (stalk-surface-above-ring):**
  - Fibrosa
  - Sedosa
  - Escamosa (con escamas)
  - Lisa
- **Superficie del Tallo por Debajo del Anillo (stalk-surface-below-ring):**
  - Fibrosa
  - Sedosa
  - Escamosa
  - Lisa
- **Color del Tallo por Encima del Anillo (stalk-color-above-ring):**
  - Marrón
  - Beige
  - Canela
  - Gris
  - Naranja
  - Rosa
  - Rojo
  - Blanco
  - Amarillo
- **Color del Tallo por Debajo del Anillo (stalk-color-below-ring):**
  - Marrón
  - Beige
  - Canela

- Gris
  - Naranja
  - Rosa
  - Rojo
  - Blanco
  - Amarillo
- 

## **Características del Velo y Anillo**

- **Color del Velo (veil-color):**
    - Marrón
    - Naranja
    - Blanco
    - Amarillo
  - **Número de Anillos (ring-number):**
    - Ninguno
    - Uno
    - Dos
  - **Tipo de Anillo (ring-type):**
    - Telaraña (algodonoso)
    - Evanescente (desaparece fácilmente)
    - Acampanado (como una falda)
    - Esquelético (fibras finas)
    - Ninguno
    - Colgante (prominente y colgante)
    - En forma de vaina
    - En zigzag
- 

## **Otras Características**

- **Población (population):**
  - Solitario (crece solo)
  - Agrupado (crece en racimos)
  - Numeroso (muchos dispersos)
  - Disperso (pocos, espaciados)
  - Extenso (crece en grupos grandes o hileras)
  - Disperso (generalmente en pares o tríos)
- **Hábitat (habitat):**
  - Bosque (crece en árboles o suelo forestal)

- Hojas (crece en hojas muertas)
- Prados (crece en campos de hierba)
- Senderos (crece en bordes de caminos)
- Urbano (crece en jardines, parques urbanos)
- Desechos (crece en madera en descomposición)
- Dunas (crece en dunas de arena)

---

**¡Importante!** Recuerda que esta aplicación es una herramienta de ayuda y **no sustituye el consejo de un experto micólogo**. Siempre consulta a un experto para la identificación final de cualquier seta que pretendas consumir. ¡Ante la duda, NO la consumas!