



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

Proyecto Final: Clasificador de hongos (Usa Aprendiziz)

Mendoza Mendoza Rosalinda
Torres Jiménez Roberto
Chavarria Juárez Marisol

Aprendizaje Maquinal.

Profesor: René Mac Kinney Romero

Data Mining a Mushroom Dataset	2
Justificación modelo.	2
Modelo J48	3
Análisis de datos	4
Gráfica atributo Olor	4
Gráfica atributo impresión de color de esporas	5
Gráfica atributo superficie del tallo	6
Gráfica atributo tipo de anillo.	8
Manual de uso.	10
Introducción	10
Objetivo	10
Requerimientos técnicos del Hardware para su ejecución.	10
Requerimientos de software	10
Funcionalidad General.	10

Data Mining a Mushroom Dataset

Justificación modelo.

Los conjunto de datos de hongos cuentan con 8124 instancias y 23 atributos, en base a esta información se podrá clasificar si un hongo es venenoso o comestible.

Claramente no existe una regla simple para determinar la comestibilidad de un hongo, por ello procedemos a utilizar los atributos más importantes.

También como parte de la práctica se dividió el conjunto de datos en dos partes, una con el 70% de datos y otra con el 30%, donde el conjunto mayor fue utilizado para aprender y el de menor tamaño fue para probarlo en el programa.

En la Tabla 1 podemos ver comparaciones del aprendiz al trabajar con todos los datos. Dado que son muchos datos el aprendizaje obtiene resultados bastante altos.

ALGORITMO	PORCENTAJE DEL APRENDIZ
Random Forest	100%
SMO	100%
DecisionTable	100%
J48	100%
Random Tree	99%
Voted Perceptron	99%
LogitBoost	98.22%

Tabla 1. Diferentes aprendizajes con 8124 instancias.

Posteriormente se realizaron los mismos aprendizajes pero con el 80% de los datos donde la precisión disminuyó en 4 de los 7 aprendizajes como se ve en la Tabla 2.

ALGORITMO	PORCENTAJE DEL APRENDIZ
Random Forest	100%
SMO	100%
DecisionTable	99.98%
J48	100%
Random Tree	99.94%
Voted Perceptron	99.77%
LogitBoost	98.19%

Tabla 2. Diferentes aprendizajes con 5318 instancias.

Modelo J48

Se ocupó el modelo j48 el cual muestra un nivel extremadamente alto de precisión que corresponde 100% de clasificación en ambas pruebas de aprendizaje.

encuentran bien clasificados, pues su barra en la gráfica tiene colores (rojo, azul) de ambas clases.

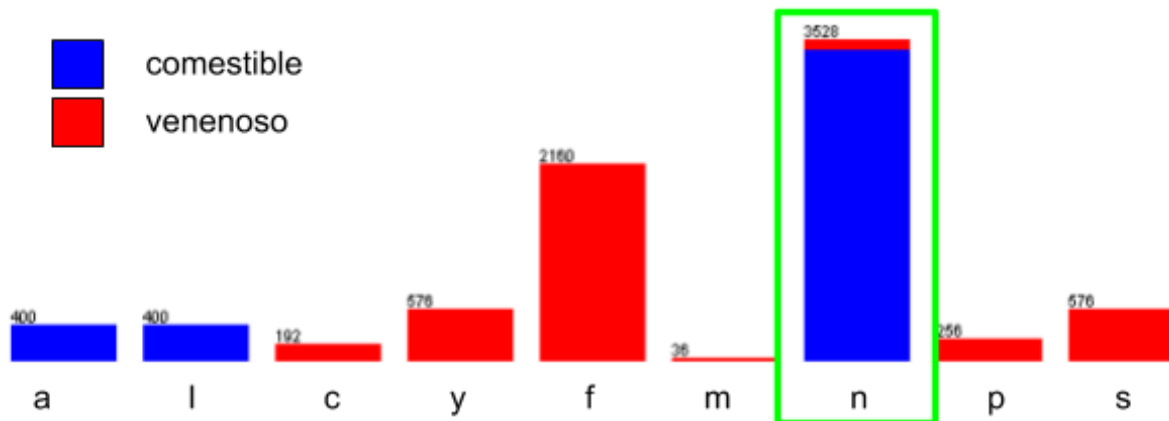


Figura 2. Distribución de clases del atributo **olor**, donde *n* comparte ambas clases.

Gráfica atributo impresión de color de esporas

Centrándonos en los datos de este valor con ambas clases y haciendo a un lado los datos que les corresponde un solo color (una sola clase), nos quedamos con 3528 datos sin clasificar como se ve en la Figura 3.

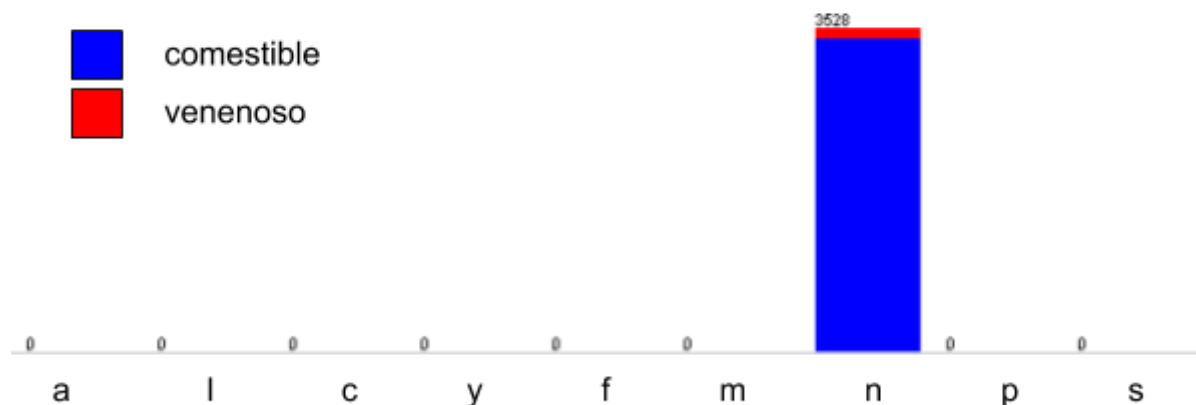


Figura 3. Instancias sin clasificar del atributo **olor** = ninguno.

Guiándonos con los atributos que sus valores permiten una clasificación casi perfecta, se tiene al atributo “**Impresión de color de esporas**” que se muestra en la Figura 4, donde 7 de 9 de sus valores dan una clasificación perfecta, un valor no tiene datos y la impresión de espora de color **amarillo** aún no tiene clasificación perfecta repitiendo el proceso que se realizó en el caso del atributo “**olor**”.

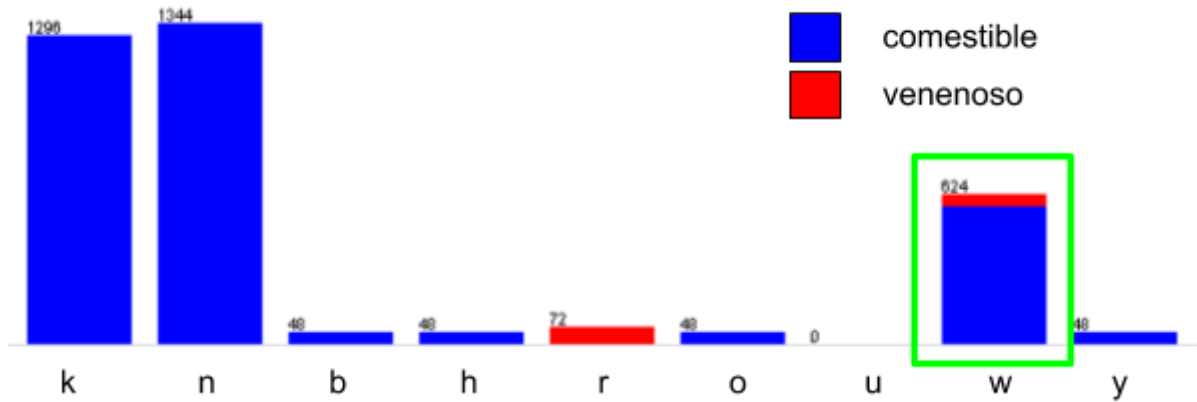


Figura 4. Distribución de clases del atributo “*impressione de color de esporas*”.

Centrándonos en los datos de este valor con ambas clases y haciendo a un lado los datos que les corresponde un solo color (una sola clase) nos quedamos con 624 datos sin clasificar como se ve en la Figura 5.



Figura 5. Instancias sin clasificar del atributo “*impressione de color de esporas*” = amarillo.

Gráfica atributo superficie del tallo

De los 624 datos faltantes por clasificar y basándonos en la casi clasificación perfecta del atributo **superficie del tallo**, notamos en la Figura 6, que solo dos barras tienen clasificación perfecta, pero diremos que una tercer barra igual da la clasificación perfecta, más adelante se da la explicación del porqué de esta decisión; en base a tener tres valores que clasifican perfecto solo nos queda un valor que no clasifica perfectamente pues la barra aun contiene las dos clases.

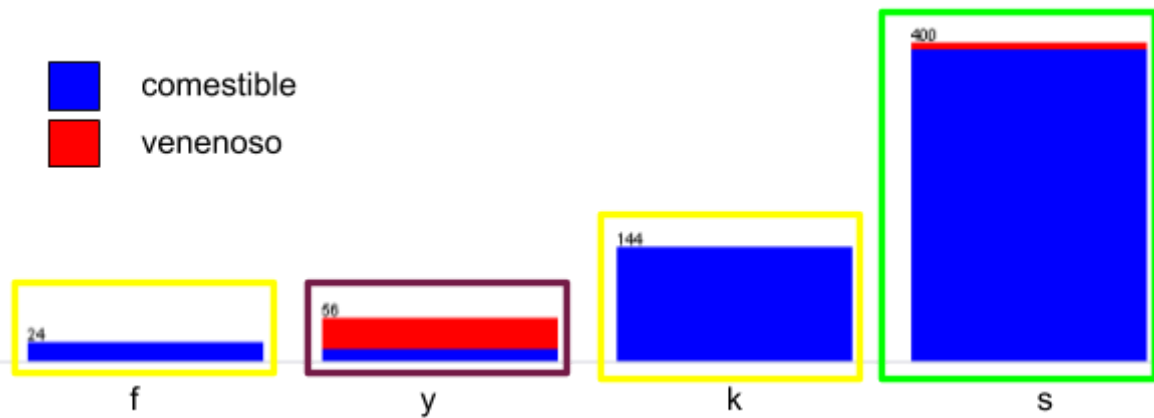


Figura 6. Distribución de clases: amarillo = perfecta, verde = se considera perfecta, morado = por clasificar.

→ **Explicación de una tercer columna con clasificación casi perfecta.**

Se decidió tomar la barra del atributo “**superficie del tallo**” con valor **liso** como si tuviera clasificación perfecta, pero como se observa en la Figura 7, tiene 400 instancias con ambas clases, donde la mayoría de estos datos son de la **clase comestible**.



Figura 7. Instancias que no clasifican perfectas con el atributo “superficie del tallo”.

Tomando estas 400 instancias y separándolos por clases como se ve en la Figura 8, observamos que sólo 8 de las 400 instancias, corresponden a la **clase venenosa**, 8 datos comparados con los 8124 totales que se tienen son una minoría por lo cual omitiremos estos datos.

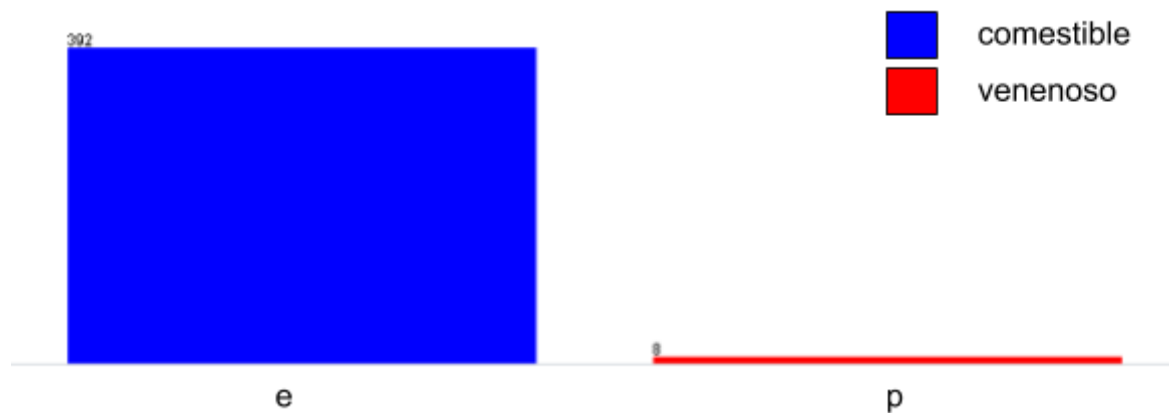


Figura 8. Solo 8 instancias pertenecen a la **clase venenosa**.

Retomando y centrándonos en la columna sin clasificar del atributo “**superficie del tallo**” con valor **escamoso**, que se muestra en la Figura 9, nos quedamos con 56 datos sin clasificar.

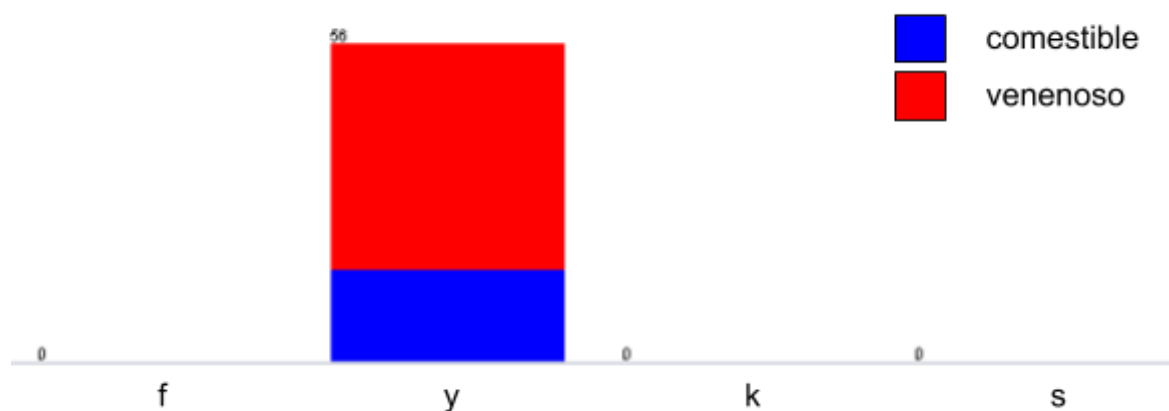


Figura 9. Instancias sin clasificar del atributo “**superficie del tallo**”.

Los 56 datos que faltan por clasificar serán clasificados con el atributo “**tipo de anillo**”.

Gráfica atributo tipo de anillo.

Como se puede observar en la Figura 10 los 56 datos son clasificados perfectamente en dos barras y a cada valor del atributo le corresponde una única clase.

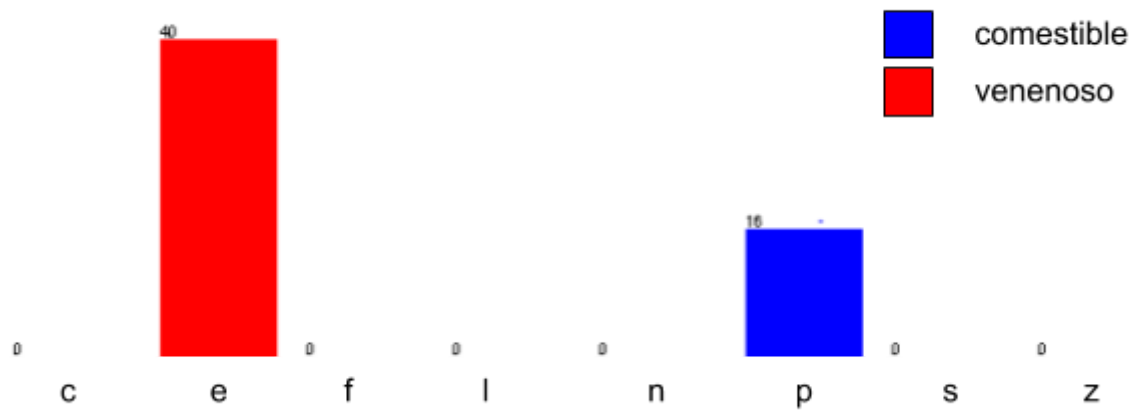


Figura 10. Distribución de clases del atributo “tipo de anillo”.

Con los resultados que se obtuvieron del análisis de datos, se tomó la decisión que los atributos que se usarán en la aplicación para preguntarle al usuario sean los siguientes.

- Olor
- Impresión de color de esporas
- Superficie del tallo
- Tipo de anillo

Manual de uso.

Introducción

El propósito de este Manual es facilitar al usuario la operación del siguiente software de clasificación sobre hongos el cual mostrará una respuesta de si dicho hongo es comestible o no comestible así como conocer en qué plataforma fue generado.

Objetivo

Clasificar los hongos que se encuentren de la mejor forma posible basándose en características (atributos) específicas sobre ellos.

Requerimientos técnicos del Hardware para su ejecución.

Contar con:

- Computadora

Requerimientos de software

Contar con:

- Sistema operativo (Windows, Linux, MacOS).
- Eclipse Jee Oxygen.
- Weka 3.8

Funcionalidad General.

- a) Una vez que se ejecuta el programa desde eclipse aparecerá una ventana como se muestra en la Figura 11, la cual tiene 5 secciones muy importantes.

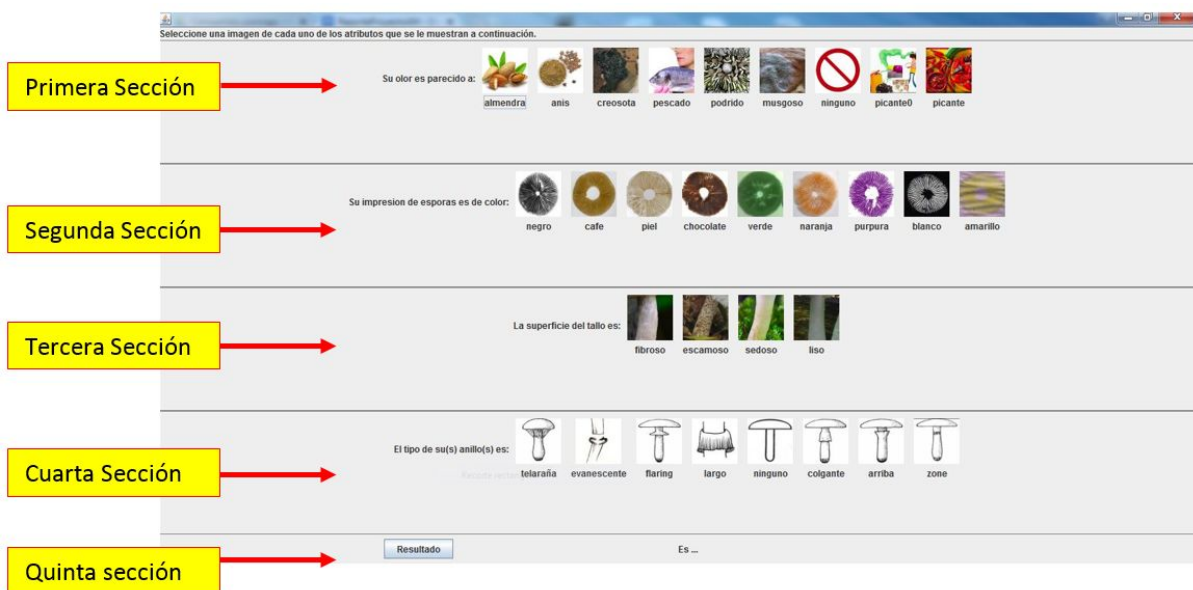


Figura 11. Pantalla principal del clasificador de hongos.

- b) En la primera sección (Figura 12), se puede elegir el olor que tiene el hongo, el cual al hacer clic sobre la respuesta que elijamos se iluminará en color verde. Podremos cambiarla cuantas veces queramos, automáticamente se le quitará la selección a la seleccionada previamente y se seleccionará la nueva.



Figura 12. Sección de selección de olor.

- c) En la segunda sección (Figura 13) se elegirá el color de su impresión de esporas al igual que en la primera sección nuestra selección se pondrá en verde.



Figura 13. Selección de impresión de color de esporas.

- d) En la tercera sección (Figura 14) elegiremos el tipo de tallo que tiene el hongo con respecto a nuestras observaciones. La respuesta seguirá resaltando en verde.

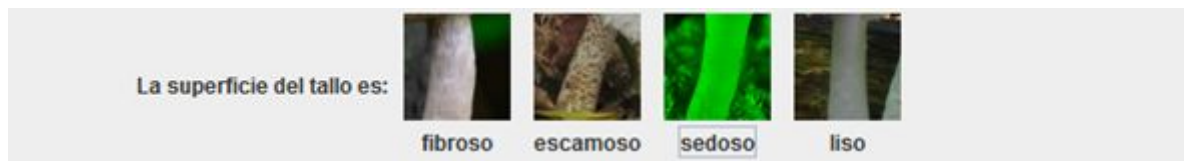


Figura 14. Selección de tipo de tallo.

- e) En la cuarta y penúltima sección elegiremos el tipo de sus anillos en el tallo si es que los tiene y si no posee también existe una respuesta que dice ninguno.



Figura 15. Selección de superficie de anillos.

- f) Y en la última sección (Figura 16) tenemos un botón que se llama “Resultado”, que al haber elegido cada uno de los atributos anteriores aplastaremos y nos mostrará en seguida una imagen sobre sí con respecto a nuestras elecciones el hongo que estemos analizando es comestible o venenoso, apareciendo una letra a su lado la cual denota comestible (Figura 16) y venenoso (Figura 17).

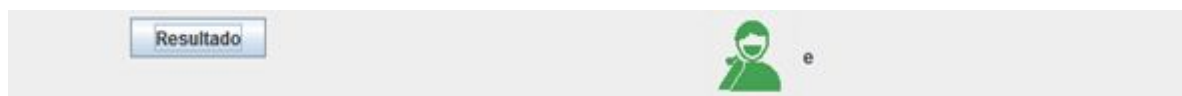


Figura 16. Sección de resultado con resultado comestible.



Figura 17. Sección de resultado con resultado venenoso.

Por último se muestran en las Figura 18 y Figura 19 se muestran las ventanas completas con una selección que resulta comestible y una que resulta venenosa respectivamente.

Seleccione una imagen de cada uno de los atributos que se le muestran a continuación.

Se oír es parecido a:

almendra anís creosota pescado podrido musgoso ninguno picante0 picante

Se impresioe de esporas es de color:

negro café piel chocolate verde naranja púrpura blanco amarillo

La superficie del tallo es:

fibroso escamoso sedoso liso

El tipo de su(s) anillo(s) es:

telaraña evanescente flaring largo ninguno colgante arriba zoe

Resultado



Figura 18. Selección con resultado comestible

Seleccione una imagen de cada uno de los atributos que se le muestran a continuación.

Se oír es parecido a:

almendra anís creosota pescado podrido musgoso ninguno picante0 picante

Se impresioe de esporas es de color:

negro café piel chocolate verde naranja púrpura blanco amarillo

La superficie del tallo es:

fibroso escamoso sedoso liso

El tipo de su(s) anillo(s) es:

telaraña evanescente flaring largo ninguno colgante arriba zoe

Resultado



Figura 18. Selección con resultado comestible

<https://www.slideshare.net/timmenzies/project-3-mushrooms>