

Clasificación de hongos por medio de aprendizaje automático

Omar D. Castrillón¹, Jaime A. Arango^{1,2} y Luis F. Castillo Ossa^{1,3}

(1) Universidad Nacional de Colombia, Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Industrial, Manizales – Colombia

(correo-e: odcastrillong@unal.edu.co, jaarangom@unal.edu.co, lfcastilloos@unal.edu.co)

(2) Universidad Católica Luis Amigó, Facultad de Ciencias Administrativas, económicas y contables, Centro Regional Manizales, Grupo de Investigación GESNE, Manizales – Colombia

(3) Universidad de Caldas, Facultad de Ingenierías, Departamento de Sistemas e Informática. Grupo de Investigación Inteligencia Artificial, Manizales – Colombia

Recibido Jun. 20, 2022; Aceptado Ago. 25, 2022; Versión final Sep. 15, 2022, Publicado Feb. 2023

Resumen

Esta investigación tiene como propósito aplicar el aprendizaje automático y la minería de datos para identificar si diferentes clases de hongos (nuevas o no) son benéficas (comestibles) o perjudiciales para el ser humano. La metodología está basada en el algoritmo de clasificación J48 (Weka) y se utiliza la base de datos online de hongos Secondary Mushroom Dataset. La variable dependiente (clase) tiene dos estados (comestible o venenoso). Se consideran 17 variables independientes, estas incluyen el hábitat, la estación del año y características morfológicas como el sombrero, los moretones o sangra, las branquias, el tallo y el anillo. Como resultado se encuentra, con una efectividad del 99.76%, que las variables más influyentes son: diámetro sombrero, superficie sombrero, color sombrero, altura tallo, grosor tallo, superficie tallo y color tallo. En conclusión, la metodología desarrollada aquí permite identificar si un hongo es venenoso o no con una efectividad del 99.76%.

Palabras clave: clasificación de hongos; aprendizaje automático; minería de datos; algoritmo J48; plataforma Weka

Classification of mushrooms through machine learning

Abstract

This research study aims to apply automated learning and data mining to determine whether different classes of mushrooms (new or not) are beneficial (edible) or harmful to humans. The methodology is based on the J48 (Weka) classification algorithm and uses the Secondary Mushroom Dataset, a fungal online database. The dependent variable (class) has two states (edible or poisonous). Seventeen independent variables are considered. They include the habitat, the season, and morphological traits such as mushroom cap, bruises or bleeds, gills, stem, ring, and class. The results show that, with an effectiveness of 99.76%, the most influential variables are: cap diameter, cap surface, cap color, stem height, stem thickness, stem surface, and stem color. In conclusion, the methodology developed here allows identifying whether a given mushroom is poisonous or not with an accuracy of 99.76%.

Keywords: mushroom classification; machine learning; data mining; J48 algorithm; Weka platform

INTRODUCCIÓN

El mercado mundial de hongos supera los 30 billones de dólares anuales, y se estima que pueden existir cerca de 2.5 Millones de especies diferentes de hongos (Cano-Estrada y Romero-Bautista, 2016). Sin embargo, solo unas cuantas, de ellas son comestibles. Dado que los hongos crecen de forma natural (silvestre) en todos los ecosistemas y sustratos de la tierra, los mismos podrían tener un gran potencial alimenticio para toda la humanidad, en especial por su valor nutricional (Cano-Estrada y Romero-Bautista, 2016). Sin embargo, para que esto sea factible debe existir una forma clara de identificar los hongos comestibles de los que no lo son, en especial cuando en los mismos cultivos de hongos pueden aparecer hongos venenosos por razones del clima, humedad, viento, etc (Rahman et al., 2022), aspecto que puede dificultar su cultivo e industrialización. Por otro lado, realizar un análisis de laboratorio para cada una de las cerca de 2.5 Millones de especies de hongos existentes resultaría por completo, inviable. En este sentido, se hace necesario una metodología de tamizaje o selección previa que permita reducir en forma considerable el número de especies de hongos que deberían ser objeto de un análisis especializado, y que además, esté al alcance de todos, con el fin de garantizar que este potencial alimenticio pueda ser aprovechado.

En consideración a lo expresado, el objetivo fundamental de este artículo es determinar por medio de la relación de 17 variables independientes (Diámetro sombrero, forma sombrero, superficie sombrero, color sombrero, moretones, branquias, espacio entre branquias, color branquias, altura tallo, grosor tallo, raíz tallo, superficie del tallo, color del tallo, anillo, clase de anillo, hábitat y estación) tomadas de una base de datos existente en la siguiente dirección web (<http://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom+Dataset>) y referenciadas en (Dua y Graff, 2019; Wagner et al., 2021) si un hongo es comestible o no, aspecto de gran trascendencia en especial cuando en algunas zonas la intoxicación por hongos se está convirtiendo en un grave problema de salud (Zhao et al., 2022). Este análisis se realiza por medio de la plataforma libre de aprendizaje automático y minería de datos denominada Weka (Witten, 2017), bajo el algoritmo J48 de comportamiento similar a los algoritmos bayesianos (Valencia et al., 2015). En el inicio, las variables independientes analizadas, son seleccionadas por medio de un proceso estadístico ejecutado bajo la plataforma Weka. Más adelante, se analiza su incidencia en la predicción del comportamiento de la variable dependiente. Como resultado de este proceso estadístico, se encuentra que con una efectividad del 99.76% las variables independientes más importantes son: Diámetro sombrero, superficie sombrero, color sombrero, altura tallo, grosor tallo, superficie tallo, color tallo.

Desde otro punto de vista, se ha encontrado que algunos hongos pueden tener grandes impactos medicinales en el tratamiento de enfermedades virales como el herpes simple, algunos coronavirus e incluso podría tener buenos resultados contra el COVID-19 (Arunachalam, et al., 2022). Estudios similares muestran que algunos hongos contienen elementos como antitumorales, antidiabéticos, antivirales, antioxidantes y aromáticos (Tian, et al., 2022), pudiendo incluso ser eficaces para disminuir la presión arterial y el colesterol en la sangre (Moumita y Das, 2022). De igual manera, también pueden ser usados en diversas áreas de la agricultura como producción y acondicionamiento (Oliveira, et al., 2022). Además, en algunos casos se ha encontrado que los hongos comestibles presentan una interacción benéfica con diferentes bacterias (Braat et al., 2022). Así mismo otros estudios (Wen y Jing, 2022) han mostrado que los hongos son ricos en cerebrosidos, moléculas con un gran valor en las funciones biológicas del cuerpo. Otros aspectos relacionados con las propiedades de los hongos pueden ser encontrados en otras referencias (Lim, et al., 2016; Xu et al., 2021).

Desde el punto de vista de la Inteligencia artificial se han empleado diversas técnicas relacionadas con los procesos de los hongos. Algunos autores (Lua y Liaw, 2020) han desarrollado, por medio de redes neuronales, un sistema de medición de imágenes con el fin de registrar el crecimiento de los mismos. También existen Sistemas de soportes decisionales para el tratamiento de intoxicaciones por hongos (Zotti et al., 2001). En este mismo sentido, otros autores han diseñado sistemas basados en lógica difusa que permiten modelar el crecimiento de los hongos (Ardabili et al., 2021). De manera análoga, se encuentran sistemas de aprendizaje profundos para la identificación de especies de hongos por medio de imágenes digitales de espectroscopia con correlación bidimensional logrando muy buenos resultados (Dong et al., 2021). Similar a este último estudio existen trabajos basados en diversas técnicas inteligentes como: Árboles de decisión, regresión logística, redes neuronales, máquinas de soporte vectorial, clasificadores bayesianos, bosques aleatorios, máquinas de aprendizaje e internet de las cosas, con el fin de identificar si un hongo es venenoso o no (Rahman et al., 2022). En esta misma área de la inteligencia artificial, existen sistemas basados en inteligencia computacional y modelado matemático para el proceso de secado de champiñones (Peter et al., 2021). En este sentido, se encuentra que los hongos también han sido usados en procesos de desalinización solar del agua (Wang et al., 2022). Para completar, se relacionan otros trabajos en diversas áreas (Schunko et al., 2022; Panda y Luyten, 2022; Saetang et al., 2002).

Como se ilustra en los párrafos anteriores, las técnicas de inteligencia artificial han sido muy poco empleadas en la identificación de los hongos. Estas técnicas solo han empezado a ser usadas en el pasado reciente,

aspecto que justifica aún más este trabajo investigativo. Por último, se expresa que con el fin de facilitar la comprensión del lector este documento ha sido organizado de la siguiente forma: a) Materiales y Métodos. En esta sección es explicada de forma detallada la metodología empleada. b) Resultados. En esta parte del se ilustran los resultados al aplicar la metodología planteada en cada uno de sus pasos. c) Discusiones. En este apartado, los resultados obtenidos en esta investigación son comparados con otros estudios similares d) Al final se incluye una breve sección de conclusiones, agradecimientos y referencias que fueron empleadas.

MATERIALES Y MÉTODOS

En el desarrollo de esta metodología, se toman como elementos previos algunos artículos de los mismos autores (Castrillón, 2021). El desarrollo de esta metodología parte de una base de datos encontrada en (<http://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom+Dataset>) y referenciada en (Dua y Graff, 2019; Wagner et al., 2021). La base de datos consta de 61069 registros y se adapta con el fin de definir cada uno de los pasos de la metodología presentada: 1) Reestructuración de la base de datos. 2) Definición del archivo Weka. 3) Selección de variables por medio de consistencia de sub conjuntos y la prueba chi cuadrado. 4) Clasificación por medio del algoritmo J48 5) Comparación con otras técnicas inteligentes de clasificación 6) Clasificación por Clúster.

Paso 1. Reestructuración de la base de datos empleada.

Se toma como referencia la base de datos mencionada en la introducción y en el párrafo anterior. Esta base de datos consta de 61.069 registros y 20 variables. No obstante, la base de datos es reestructurada con el fin de dejar solo aquellas variables que tengan información completa en todos sus campos o que sea posible completarlas con un indicativo especial (NE). En tal sentido se seleccionan 18 variables que cumplen con esta condición. 17 de las variables seleccionadas son independientes y una variable dependiente. La estructura de todos los campos y tipos de datos de la nueva base de datos definida para este artículo es ilustrada en la Tabla 1.

Tabla 1: Estructura de la base de datos empleada

#	Variable	Abrev	Unidades de medida / valores	Tipo Variable
1	Diámetro sombrero	T_som	Cms	Independiente
2	Forma sombrero	F_som	b=campana, c=conica, x=convexa, f= plana, s=hundida, p= esférica, o= otras, NE = dato en blanco.	Independiente
3	Superficie sombrero	S_som	i=fibroso, g=surcos, y=escamosa, s= lisa, h=brillante, l=correosa, k=sedosa, t=pegajosa, w=arrugada, e=carnosa, d= otra, NE=dato en blanco	Independiente
4	Color sombrero	C_som	n=marrón, b= brillante, g=gris, r= verde, p=rosado, u=purpura, e=rojo, w=blanco, y=amarillo, l= azul, o=naranja, k=negro, NE=dato en blanco	Independiente
5	moretones o sangra	M_s	t=verdadero, f=falso	Independiente
6	Branquias	B	a=adnato, x=anexo, d=decurrente, e=libre, s=ondeado, p=poros, f=ninguno, ?=desconocido, NE = dato en blanco.	Independiente
7	Espacio branquias	B_e	c=estrecho, d=distante, f=ninguno, NE = dato en blanco.	Independiente
8	Color branquias	B_c	n=marrón, b= amarillo piel, g=gris, r= verde, p=rosado, u=purpura, e=rojo, w=blanco, y=amarillo, l= azul, o=naranja, k=negro, f=otro, NE = dato en blanco.	Independiente
9	Altura tallo	A_t	Cms	Independiente
10	Grosor tallo	G_t	Cms	Independiente
11	Raíz del tallo	R_t	b=bulbosa, s=hinchada, c= agrupada, u=sombrero, e=igual, z=rizomorfos, r= arraigada, f=otro, NE = dato en blanco	Independiente
12	Superficie del tallo	S_t	i=fibroso, g=surcos, y=escamosa, s= lisa, h=brillante, l=correosa, k=sedosa, t=pegajosa, f=otro, NE = dato en blanco	Independiente
13	Color del tallo	C_t	n=marrón, b= brillante, g=gris, r= verde, p=rosado, u=purpura, e=rojo, w=blanco, y=amarillo, l= azul, o=naranja, k=negro, f=otro, NE=dato en blanco	Independiente
14	Anillo	A	t=verdadero, f= falso	Independiente
15	Clase anillo	A_c	c=telarañas, e=evanescente, r=resplandeciente, g=estriado, l=largo, p=colgante, s=revestido, z=zona, y=escamoso, m=móvil, f=otro, ?=desconocido, NE = dato en blanco	Independiente
16	Habitat	H	g=pastos, m=prados, p=caminos, h=brezales, u=urbano, w=basuras, d=bosques, l=hojas, NE = dato en blanco	Independiente
17	Estación	E	s=primavera, u=verano, a=otoño, w=invierno, NE = dato en blanco	Independiente
18	Clase	C	e=comestible, p=venenoso	Dependiente

Paso 2. Definición del archivo Weka

Con el fin de analizar la base de datos creada en el paso anterior, se definen el respectivo archivo .Arff compuesto de dos partes denominadas cabeza y detalle. La definición de este archivo permite por medio de la plataforma libre de aprendizaje automático y minería de datos denominada Weka, seleccionar las variables más influyentes en el proceso de clasificación, empleando la prueba de consistencia de subconjuntos y la prueba chi cuadrado de la misma plataforma. Además, por medio del algoritmo J48 y el método de clúster se realiza la respectiva predicción de la variable dependiente, como se ilustra en los dos pasos siguientes.

Paso 3. Selección de variables.

Esta opción denominada *ConsistencySubSetEval* en la plataforma weka, permite seleccionar el menor de todos los subconjuntos de variables independientes, con un rendimiento igual al rendimiento completo de todas las variables independientes. Esta técnica evita la pérdida de información al momento de realizar la predicción de la variable dependiente. Posterior a este proceso se realiza un ranqueo de las variables seleccionadas por medio de la prueba Chi-Cuadrado. Este ranqueo, permite seleccionar las variables que revistan por lo menos el 80% de la información necesaria para la predicción de la variable dependiente.

Paso 4. Clasificación por medio del algoritmo J48.

Los atributos o variables independientes, seleccionadas en el proceso anterior, son empleados para realizar el proceso de clasificación en la plataforma Weka, por medio del algoritmo J48. Versión libre del algoritmo C4.5, el cual permite realizar sus clasificaciones mediante el manejo de la entropía de la información. Este algoritmo puede ser empleado en todas las áreas (medicina, economía, biología, matemáticas, física, etc.) en problemas de relacionados con la predicción de variables dependientes o clusterización de los datos. Este proceso de clasificación se realiza por medio de tres parametrizaciones del algoritmo J48: 1) Entrenando y validando con el 100% de los datos. 2) Entrenando con el 50% y 75% de los datos y validando con el 50% y 25% de los datos respectivamente 3) Por medio de una validación cruzada con los siguientes porcentajes de entrenamiento y validación: 70% - 30%, 80% - 20%, 90%-10%. Luego, se definen las respectivas matrices de clasificación, precisión, confusión, y las relaciones entre las variables independientes y la variable dependiente. Es importante evitar los sesgos que se puedan producir debido a un sobre entrenamiento del sistema. Este último aspecto no permitiría generalizar las reglas, evitando realizar una adecuada predicción de la variable dependiente.

Paso 5. Optimización por medio de variables calculadas.

Se busca, la mejor relación entre las posibles variables numéricas que hallan sido seleccionadas en este proceso. Esta relación es agregada como variable calculada al archivo Arff construido, con el fin de mejorar el proceso de predicción de la variable dependiente.

Paso 6. Comparación con otras técnicas inteligentes de clasificación.

Con el fin de establecer la efectividad de las técnicas de clasificación empleadas, se realiza una comparación frente a otras técnicas inteligentes de clasificación, bajo la misma plataforma de aprendizaje automático y minería de datos denominada Weka. Esta comparación se realiza por medio de una validación cruzada 80% - 20% con técnicas de: Bayes, regresión, lazy, reglas y árboles.

Paso 7. Clasificación por clústeres.

En primer lugar, se emplea el algoritmo *expectation maximisation (EM)* de la plataforma weka, con el fin de establecer el número de grupos de interés según sus probabilidades. Establecido el número de grupos de interés, se emplea el algoritmo SimpleKMeans de la plataforma Weka para generar los diferentes grupos de interés, con el fin de definir cuál o cuáles de ellos pueden ser clasificados en cada uno de los valores de la variable dependiente, así como su porcentaje de efectividad. Dado que la variable dependiente solo presenta dos estados, cada grupo de interés deberá pertenecer a un estado u otro estado de la variable dependiente. Aspecto que permite su clasificación.

RESULTADOS

Aplicada la metodología definida en la sección anterior en cada uno de sus pasos, se encuentran los resultados, que se ilustran a continuación en esta parte de la investigación, detallados paso por paso y conservando el mismo orden que se empleó en la sección de Materiales y Métodos. Los siguientes párrafos de esta ilustran este aspecto :

Paso 1. Restructuración de la base de datos empleada

Como resultado, en este punto de la metodología se define una base de datos con 61.069 registros y 18 variables con sus respectivos nombres abreviados y adaptados al español, como se ilustró en la Tabla 1. En forma similar, la Tabla 2, muestra un fragmento de esta base de datos que es una adaptación de la base de datos encontrada en la página : (<http://archive.ics.uci.edu/ml/datasets/Secondary+Mushroom+Dataset>) y referenciada en (Dua y Graff, 2019; Wagner et al., 2021). Por razones de espacio solo se ilustra una fracción de esta Base de Datos.

Tabla 2: Fragmento base de datos adaptada.

<i>T_som</i>	<i>F_som</i>	<i>S_som</i>	<i>C_som</i>	<i>M_s</i>	<i>B</i>	<i>B_e</i>	<i>B_c</i>	<i>A_t</i>	<i>G_t</i>	<i>R_t</i>	<i>S_t</i>	<i>C_t</i>	<i>A</i>	<i>A_c</i>	<i>H</i>	<i>E</i>	<i>C</i>
7.63	f	s	n	f	s	c	w	5.74	10.14	b	i	n	f	f	d	a	e
7.04	f	s	n	f	s	c	w	6.29	10.23	b	i	n	f	f	d	a	e
6.65	f	s	n	f	s	c	w	6.89	10.67	b	i	n	f	f	d	a	e
7.22	f	s	n	f	s	c	w	6.83	11.55	b	i	n	f	f	d	a	e
8.63	f	s	n	f	s	c	w	7.22	11.6	b	i	n	f	f	d	a	e

Paso 2. Definición del archivo Weka

Como resultado de este paso de la metodología, se obtienen las Tablas 3 y 4, donde se ilustra la estructura del archivo Arff, base fundamental para el desarrollo de las siguientes secciones de la metodología, bajo la plataforma Weka.

Tabla 3: Encabezado archivo Arff

Relación	Variable	Tipo
@relation	Hongos	
@attribute	T_som	numeric
@attribute	F_som	{b,c,x,f,s,p,o,NE}
@attribute	S_som	{i, g, y, s, h,l,k,t,o,d,w,e,NE}
@attribute	C_som	{n,b,g,r,p,u,e,w,y,l,o,k, NE}
@attribute	M_s	{f,t}
@attribute	B	{a,x,d,e,s,p,f,NE}
@attribute	B_e	{c,d,f, NE}
@attribute	B_c	{n,b,g,r,p,u,e,w,y,l,o,k,f,NE}
@attribute	A_t	numeric
@attribute	G_t	numeric
@attribute	R_t	{b, s, c, u,e,z,r,f,NE}
@attribute	S_t	{i,g,y,s,h,l,k,t,f,NE}
@attribute	C_t	{n,b,g,r,p,u,e,w,y,l,o,k,f,NE}
@attribute	A	{t,f}
@attribute	A_C	{c, e, r, g,l,p,s,z,y,m,f,?,NE}
@attribute	H	{g,m,p,h,u,w,d,l,NE}
@attribute	E	{s,u,a,w,NE}
@attribute	C	{e,p}

Tabla 4: Detalle Archivo Arff

<i>T_som</i>	<i>F_som</i>	<i>S_som</i>	<i>C_som</i>	<i>M_s</i>	<i>B</i>	<i>B_e</i>	<i>B_c</i>	<i>A_t</i>	<i>G_t</i>	<i>R_t</i>	<i>S_t</i>	<i>C_t</i>	<i>A</i>	<i>A_c</i>	<i>H</i>	<i>E</i>	<i>C</i>
4.98	b	d	n	f	a	c	n	5.1	5.84	NE	i	n	t	NE	d	a	p
5.85	b	d	n	f	a	c	y	5.25	6.15	NE	i	y	t	NE	d	a	p
5.79	b	d	n	f	a	c	y	5.21	6.44	NE	i	n	t	NE	d	a	p
5.55	b	d	n	f	d	c	n	5.48	6.56	NE	i	y	t	NE	d	u	p
5.55	b	d	n	f	a	c	n	5.13	6.66	NE	i	n	t	NE	d	u	p
6.09	b	d	n	f	d	c	n	5.47	6.56	NE	i	n	t	NE	d	a	p
6.46	b	d	n	f	a	c	y	6.1	7.17	NE	i	y	t	NE	d	u	p

Paso 3. Selección de variables

Una clasificación inicial de la variable dependiente del archivo Arff construido en el paso anterior, por medio del algoritmo J48 bajo la plataforma Weka establece un nivel de éxito del 99.87% en la predicción de esta variable. No obstante, por medio la técnica denominada “*ConsistencySubSetVal*” de la misma plataforma Weka, se selecciona el menor subconjunto de variables independientes con un rendimiento igual al conjunto total de variables independientes. En este caso, las variables independientes seleccionadas son: T_som, S_som, C_som, B, B_e, B_c, A_t, G_t, S_t, C_t. Con las anteriores variables y por medio del algoritmo J48, se logra un nivel de efectividad del 99.91% en la predicción de la variable dependiente. A continuación, por

medio de la técnica Chi – Cuadrado, bajo la misma plataforma, se realiza un ranqueo de las variables preseleccionadas, escogiendo aquellas que por lo menos contengan el 80% de la información como se muestra en la Tabla 5. En la Tabla 5, se evidencia que los 7 primeros atributos (señalados con ***), representan el 83.18% del peso total de todas las variables independientes. En tal sentido, los atributos seleccionados, con los que se continuará el proceso de clasificación son: G_t, C_t, A_t, S_t, T_som, C_som, S_som.

Tabla 5: Ranking de las variables

Variable	Peso
G_t	6906,5974***
C_t	4557,8906***
A_t	4552,9497***
S_t	4045,5581***
T_som	3581,2771***
C_som	3541,0798***
S_som	3188,13 ***
B	3094,3009
B_c	2245,0489
B_e	799,8579

Paso 4. Clasificación por medio del algoritmo J48.

Con base en las 7 variables independientes seleccionadas en el paso anterior, se realiza un nuevo proceso de clasificación por medio del mismo algoritmo J48. Como resultado de este proceso se predice la variable dependiente con una efectividad del 99.29% por medio de una validación cruzada 90% - 10%. Las Tablas 6, 7 y 8 muestran las respectivas matrices clasificación, precisión, confusión. la Tabla 9, ilustra los diferentes porcentajes de clasificación logrados según los porcentajes de entrenamiento y validación. Así mismo las Tablas 10 y 11 muestran las principales relaciones entre variables independientes para los hongos comestibles y venenosos. Las Tablas 10 y 11 son equivalentes al árbol de clasificación generado desde weka. Se descartan las relaciones con el atributo NE (blanco Tabla 1) para evitar sesgos.

Tabla 6: Matriz de clasificación

Instancias	#	%
Instancias clasificadas en forma correcta	60637	99.2926 %
Instancias clasificadas en forma incorrecta	432	0.7074 %
Estadística Kappa	0.9857	
Error absoluto medio	0.0095	
Error cuadrático medio	0.082	
Error absoluto relativo	1.915 %	
Error cuadrático relativo	16.5069 %	
Número de instancias totales	61069	

Tabla 7: Matriz de precisión

Positi verdaderos	Falsos Positivos	Precision	Re calculo	Medida F	Area ROC	Clase
0.992	0.006	0.992	0.992	0.992	0.996	e
0.994	0.008	0.994	0.994	0.994	0.996	p
0.993	0.007	0.993	0.993	0.993	0.996	Promedio

Tabla 8: Matriz de confusión

a	b	Clasificación
26964	217	a = e
215	33673	b = p

Tabla 9: Porcentajes de clasificación logrados

Método	% Entrenamiento	% Validación	% Éxito
Set Completo	100%	100%	99.70
Cruzada	90%	10%	99.29
Partición	66%	33%	99.24%
Partición	50%	50%	99.17%
		Promedio →	99.35%

Tabla 10: Relaciones entre variables independientes. Hongos comestibles.

<i>Relaciones variables independientes hongos comestibles</i>
$S_t=i \rightarrow S_som=(g, y)$
$S_t=i \rightarrow S_som=s \rightarrow G_t \leq 11.72$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=11.72 \text{ y } <= 15.28) \rightarrow A_t (>6.31 \text{ y } <=7.92) \rightarrow C_som = (n, b, p, u, e, w, y, l, o, k)$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=11.72 \text{ y } <= 17.06) \rightarrow A_t >7.92$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=17.06) \rightarrow A_t > 9.15$
$S_t=i \rightarrow S_som=h \rightarrow G_t \leq 15.11$
$S_t=i \rightarrow S_som=t \rightarrow t_som > 5.5$
$S_t=y \rightarrow S_som=i$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t=n \rightarrow A_t \leq 7.82 \text{ ó } >=7.68 \rightarrow G_t >5.97$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t= (b, g, r, p, u, e, w, l, o, k, f,)$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som >5.07$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t \leq 4.92 \rightarrow G_t \leq 5.33 \rightarrow T_som >4.07$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 4.33 \rightarrow A_t \leq 4.92 \rightarrow G_t >5.33$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t >4.63 \text{ y } <= 4.92 \rightarrow G_t \leq 5.88 \rightarrow T_som >4.43$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t >4.92 \leq 5.23 \rightarrow G_t \leq 6.12$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=b, g, r, p, u, e, w, y, l, k, f)$
$S_t=y \rightarrow S_som=d$
$S_t=s \rightarrow T_som > 2.09 \text{ y } <= 4.06 \rightarrow S_som = g$
$S_t=s \rightarrow T_som \leq 4.06 \rightarrow S_som = (s, h, t)$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = (l, g)$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = y \rightarrow C_som = w$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = s \rightarrow C_t = (n, b, g, r, p, u, e)$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = s \rightarrow C_t = w \rightarrow C_som = (b, r, p, u, e, w, y, l, o, k)$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = s \rightarrow C_t = (y, l, o, k, f)$
$S_t=s \rightarrow T_som >= 4.06 \rightarrow S_som = (h, l, k, t, o, e)$
$S_t = k \rightarrow T_som \leq 30.94 \rightarrow S_som = (h, t)$
$S_t = k \rightarrow T_som > 30.94$
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = w$
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = g \rightarrow G_t \leq 8.22$
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = y \rightarrow G_t \leq 2.08$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 3.52 \rightarrow C_som = (n, b, g, u, e, w, l, o, k)$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som \leq 1.73 \rightarrow C_som = n \rightarrow S_som = (i, g, y, s, l, k, o, d, w, e)$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som \leq 1.73 \rightarrow C_som = n \rightarrow S_som = t \rightarrow G_t \leq 2.06$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som \leq 1.73 \rightarrow C_som = (b, g, r, p, u, e, w, y, l, o, k)$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = (o, w)$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = y \rightarrow S_som = (g, s)$
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = y \rightarrow S_som = t \rightarrow T_som \leq 1.87 \rightarrow C_som = y$
$S_t = f$

Tabla 11: Relaciones entre variables independientes. Hongos venenosos.

<i>Relaciones variables independientes hongos venenosos</i>
$S_t=i \rightarrow S_som=i$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=11.72 \text{ y } <= 15.28) \rightarrow A_t (>6.31 \text{ y } <=7.92) \rightarrow C_som = (g, r)$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=11.72 \text{ y } <= 15.28) \rightarrow A_t \leq 6.31$
$S_t=i \rightarrow S_som=s \rightarrow G_t > 15.28$
$S_t=i \rightarrow S_som=s \rightarrow G_t (>=17.06) \rightarrow A_t \leq 9.15$
$S_t=i \rightarrow S_som=h \rightarrow G_t > 15.11$
$S_t=i \rightarrow S_som=(l, k)$
$S_t=i \rightarrow S_som=t \rightarrow t_som \leq 5.5$
$S_t=i \rightarrow S_som=(o, d, w, e)$
$S_t=g$
$S_t=y \rightarrow S_som=g$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t=n \rightarrow A_t > 7.82$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t=n \rightarrow A_t \leq 7.82 \rightarrow G_t \leq 5.97$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t=n \rightarrow A_t \leq 7.52 \text{ ó } >=7.68 \rightarrow G_t >5.97$
$S_t=y \rightarrow S_som=y \rightarrow C_som=n \rightarrow C_t=y$
$S_t=y \rightarrow S_som=y \rightarrow C_som=(b, g, r, p, u, e, w, y, l, o, k)$
$S_t=y \rightarrow S_som=s$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t \leq 4.92 \rightarrow G_t \leq 5.88 \rightarrow T_som \leq 4.07$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t \leq 4.63 \rightarrow G_t \leq 5.88 \rightarrow T_som >4.43$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t \leq 4.92 \rightarrow G_t >5.88$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t >5.23 \text{ y } <= 5.42 \rightarrow G_t > 6.12$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=n \rightarrow T_som \leq 5.07 \rightarrow A_t >5.23 \rightarrow G_t > 6.12$
$S_t=y \rightarrow S_som=h \rightarrow C_som =n \rightarrow c_t=o$

Tabla 11: continuación

<i>Relaciones variables independientes hongos venenosos</i>	
$S_t=y \rightarrow S_som=h \rightarrow C_som = (b, g, r, p, u, e, w, y, l, o, k)$	
$S_t=y \rightarrow S_som = (l, k, t, o, w, e)$	
$S_t=s \rightarrow T_som \leq 4.06 \rightarrow S_som = (l, y, l, k, o, d, w, e)$	
$S_t=s \rightarrow T_som \leq 2.09 \rightarrow S_som = g$	
$S_t=s \rightarrow T_som \geq 4.06 \rightarrow S_som = y \rightarrow C_som = (n, b, g, r, p, u, e, y, l, o, k)$	
$S_t=s \rightarrow T_som \geq 4.06 \rightarrow S_som = s \rightarrow C_t = w \rightarrow C_som = (g, n)$	
$S_t=s \rightarrow T_som \geq 4.06 \rightarrow S_som = (d, w)$	
$S_t = (h, l)$	
$S_t = k \rightarrow T_som \leq 30.94 \rightarrow S_som = (l, g, y, s, l, k, o, d, w, e)$	
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = (e, u, p, b, n, r, l, o, k, f)$	
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = g \rightarrow G_t > 8.22$	
$S_t = t \rightarrow A_t \leq 5.2 \rightarrow C_t = y \rightarrow G_t > 2.08$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 3.52 \rightarrow C_som = (r, p, y)$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som \leq 1.73 \rightarrow C_som = n \rightarrow S_som = h$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som \leq 1.73 \rightarrow C_som = n \rightarrow S_som = t \rightarrow G_t > 2.06$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = (n, b, g, r, p, u, e, l, k, f)$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = y \rightarrow S_som = (l, y, h, i, k, o, d, w, e)$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = y \rightarrow S_som = t \rightarrow T_som > 1.87$	
$S_t = t \rightarrow A_t > 5.2 \rightarrow T_som > 1.73 \rightarrow C_t = y \rightarrow S_som = t \rightarrow T_som \leq 1.87 \rightarrow C_som = (n, b, g, r, p, u, e, w, o, k)$	

Paso 5. Optimización por medio de variables calculadas.

Los porcentajes de clasificación obtenidos en la Tabla 9, son optimizados por medio de una variable calculada, en tal sentido, se buscan todas las relaciones posibles entre las variables numéricas que fueron seleccionadas en el paso 5 de esta sección. Estas variables son: T_som , A_t , G_t . La Tabla 12, ilustra los nuevos porcentajes de clasificación, al adicionar la posible relación de todas las variables calculadas, al entrenar y validar el sistema con todo el conjunto de datos. Encontrándose la relación A_t/G_t como la mejor de ellas.

Tabla 12: Variables calculadas

A_i/A_j	T_som	A_t	G_t
T_som		99.73%	99.70%
A_t	99.73%		99.76%***
G_t	99.70%	99.76%***	

Paso 6. Comparación con otras técnicas inteligentes de clasificación.

Los resultados obtenidos en el paso 5, de esta sección son comparados con otras técnicas de clasificación basadas en inteligencia artificial, por medio de un proceso de validación cruzada 90% -10%. Con 10 folds, 10 réplicas bajo la métrica de desempeño Accuracy. La Tabla 13, ilustra esta comparación entre diferentes técnicas de inteligencia artificial encontradas en la misma plataforma denominada Weka. Además, en la columna denominada “Todas las variables” de esta misma tabla se establece la efectividad al entrenar y validar con el 100% de los datos y las 18 variables iniciales.

Tabla 13: % Éxito al comparar con otras técnicas de Inteligencia artificial. Parámetros por defecto del Weka.

Técnica	Con 7 variables	Todas las variables.
Bayes Net	74.75%	78.11%
Logistic	74.38%	85.10%
RBFNetwork	71.36%	77.89%
IB1	99.60%	100%
IBK	99.60%	100%
Decisión Table	96.11%	99.58%
J48Graf	99.28%	99.97%
RandomForest	99.52%	100%
Random Tree	99.19%	100%

Paso 7. Clasificación por clústeres.

Al comienzo, por medio del algoritmo EM, de la plataforma de Minería de datos y aprendizaje automática denominada Weka, se logran determinar 13 grupos de interés, así como la cantidad de datos con sus respectivos porcentajes que pertenecen a cada grupo. Ver Tabla 14. Más adelante, por medio del algoritmo SimpleKmeans, de esta misma plataforma, se generan cada uno de los grupos de interés establecidos por el algoritmo EM. Las Tablas 15 y 16, ilustran este aspecto.

Tabla 14: Grupos de interés

Grupo	Cantidad	Porcentaje
0	2280	4%
1	7026	12%
2	2013	3%
3	2623	4%
4	2811	5%
5	11646	19%
6	9119	15%
7	3525	6%
8	4650	8%
9	6107	10%
10	4623	8%
11	2833	5%
12	1813	3%
Total	61.069	100%

Tabla 15: Características de los grupos de interés. Grupos 0 - 5.

Grupos →	Todo	0	1	2	3	4	5
	61069	2280	7026	2013	2623	2811	11646
T_som	6.7339	3.7363	6.0643	5.1843	5.032	12.8172	7.9462
S_som	NE	t	NE	y	i	NE	NE
C_som	n	n	w	o	e	y	n
A_t	6.5815	5.8777	6.0224	6.0739	5.923	6.511	6.4837
G_t	12.1494	5.5882	11.1656	9.971	7.1453	17.3074	14.672
S_t	NE	t	NE	NE	NE	NE	NE
C_t	w	y	w	o	e	y	w
C	p	e	p	p	p	e	e

Tabla 16: Características de los grupos de interés. Grupos 6 - 12.

Grupos →	6	7	8	9	10	11	12
	9119	3525	4650	6107	4623	2833	1813
T_som	9.2716	3.5421	6.1747	3.1883	4.9752	6.7066	11.4069
S_som	t	s	h	g	s	t	NE
C_som	n	g	n	n	y	e	n
A_t	7.78	3.3703	6.8788	5.2232	5.4102	9.8188	13.8445
G_t	20.5703	5.1941	9.9482	4.2364	9.3686	11.6859	20.9538
S_t	NE	NE	s	s	NE	NE	y
C_t	n	w	n	n	y	w	w
C	p	p	e	p	p	p	e

DISCUSIÓN

Un análisis de las Tablas 10 y 11 asocia la variable S_t, (Superficie del tallo) como la variable principal que permite la identificación de los hongos. En este sentido se encuentra que un hongo es comestible cuando se dan los siguientes factores:

Primero. La superficie del tallo es fibrosa, y además se cumple alguno de los siguientes literales: a) La superficie del sombrero del hongo es en surcos o escamosa. b) la superficie del sombrero es lisa, con un grosor de tallo menor a 11.72 cm. c) la superficie del sombrero es lisa y el grosor del tallo esta entre 11.72 cms y 15.28 cms con una altura entre 6.31 cms y 7.92 cms y color de sombrero: marrón, brillante, rosado, púrpura, rojo, blanco, amarillo, azul, naranja o negro. c) la superficie del sombrero es lisa, y el grosor del tallo es mayor a 17.06 y su altura mayor a 9.15 cms. d) la superficie del sombrero es brillante y el grosor del tallo es menor a 15.11 cms. e) la superficie del sombrero es pegajosa y su diámetro mayor a 5.5 cms.

Segundo. La superficie del tallo es escamosa, y además se cumple al menos una de las siguientes condiciones: a) la superficie del sombrero es fibrosa. b) la superficie del sombrero es escamosa, el color del sombrero es marrón, el color del tallo brillante, gris, verde, rosado, púrpura, rojo, blanco, azul, naranja o negro. En este caso si el color del tallo es marrón la altura del tallo deberá estar entre 7.82 y 7.68 con un grosor de tallo mayor a 5.97 cms. c) la superficie del sombrero es brillante, de color marrón, al igual que su

tallo con un diámetro inferior a 4.03 cms y una altura y grosor de tallo proporcionales a este tamaño. d) la superficie del sombrero es brillante, de color marrón, al igual que su tallo con un diámetro superior a 5.07 cms. e) la superficie del sombrero es brillante, de color marrón, y el tallo es de color: brillante, gris, verde, rosado, púrpura, rojo, blanco, amarillo, azul o negro.

Tercero. La superficie del tallo es lisa, el hongo es comestible en los siguientes casos: a) si el diámetro del sombrero es mayor a 2.09 cms y menor a 4.06 cms con una superficie de sombrero en forma de surcos. b) Si el diámetro del sombrero es menor a 4.06 cms y la superficie del sombrero es lisa, brillante o pegajosa. c) si el diámetro del sombrero es mayor que 4.06 cms, la superficie del sombrero es brillante, correosa, sedosa, pegajosa, carnosa, surcos. En los casos que la superficie del sombrero es lisa, los hongos son comestibles si el color del tallo no es blanco. En el caso anterior, cuando el color del tallo es blanco, es comestible si el color del sombrero es brillante, verde, rosado, púrpura, rojo, blanco, amarillo, azul, naranja o negro. d) Si el diámetro del sombrero es mayor a 4.06 cms y la superficie del sombrero es brillante, correosa, sedosa, pegajosa o carnosa.

Cuarto. La superficie del tallo es sedosa, o pegajosa. En el primer caso (sedosa) el hongo es comestible si el diámetro del sombrero es mayor que 30.94 cms. Cuando el diámetro del sombrero tiene un valor menor, el hongo es comestible si su superficie es brillante o pegajosa. En el segundo caso (pegajosa), el hongo es comestible en los siguientes casos: a) el ancho del tallo es menor que 5.2 cms, y además el color del tallo es blanco. b) el ancho del tallo es menor que 5.2 cms y el color del tallo es gris o amarillo con un ancho de tallo inferior a 8.22 o 2.08 cms respectivamente. c) el ancho del tallo es mayor a 5.2 cms y el tamaño del sombrero es mayor a 3.52 cms y el color del sombrero es marrón, brillante, gris, púrpura, rojo, blanco, azul, naranja o negro. d) el ancho del tallo es mayor a 5.2 cms, el tamaño del sombrero es menor que 1.73 cms es comestible si el color del sombrero no es marrón. Cuando el color del sombrero es marrón, será comestible si la superficie del sombrero es fibrosa, surca, escamosa, lisa, correosa, sedosa, arrugada o carnosa. Cuando la superficie del sombrero es pegajosa dependerá que el ancho del tallo no sea mayor a 2.06 cms. e) el ancho del tallo es mayor a 5.2 cms y el tamaño del sombrero es mayor que 1.73 cms y un color de tallo naranja, blanco. Cuando el color del tallo es amarillo, dependerá que la superficie del sombrero sea en surcos o lisa. Si la superficie del sombrero es pegajosa y el color del sombrero es amarillo, el hongo también es comestible.

En un sentido contrario a lo expresado atrás, se puede decir que *un hongo es venenoso* cuando se tienen las siguientes situaciones:

Primero. La superficie de tallo es fibrosa, se puede establecer que el hongo es venenoso en los siguientes casos: a) si la superficie del sombrero es fibrosa, correosa o sedosa b) si la superficie del sombrero es brillante o lisa y el grosor del tallo es mayor que 15.28 cms. c) si la superficie del sombrero es lisa y el grosor del tallo es mayor que 17.06 con una altura de tallo mayor que 9.15 cms. d) si el grosor del tallo esta entre 11.72 y 15.28 cms y la altura del tallo es menor que 6.31 cms. Si la altura esta entre 6.31 y 7.92 cms, el color del sombrero debe ser gris o verde. e) si la superficie del sombrero es carnosa, arrugada o pegajosa, en este último caso (pegajosa) el diámetro de su sombrero debe ser menor que 5.5 cms.

Segundo. La superficie del tallo es en surcos, brillante o correosa el hongo es venenoso, o si La superficie del tallo es escamosa, el hongo también es venenoso en los siguientes casos: a) Si la superficie del sombrero es en surcos, lisa, correosa, sedosa, pegajosa, arrugada o carnosa b) si la superficie del sombrero es escamosa, el color del sombrero y tallo es marrón, en este caso la altura de tallo debe ser superior a 7.82 cms. Si su altura no es superior a este valor, su grosor de tallo deberá ser inferior a 5.97 cms c) si la superficie del sombrero es escamosa, el color del sombrero marrón y el color del tallo amarillo d) si la superficie del sombrero es escamosa, el color del sombrero es brillante, gris, verde, rosado, púrpura, rojo, blanco, amarillo, azul, naranja o negro. e) si la superficie del sombrero es brillante y el color del sombrero es brillante, gris, verde, rosado, púrpura, rojo, blanco, amarillo, azul, naranja o negro. En este último caso si el color del sombrero es marrón el color del tallo debe ser naranja. f) si la superficie del sombrero es brillante, el color del sombrero y el del tallo marrón, con un diámetro menor que 5.07 cms.

Tercero. La superficie del tallo es lisa, el hongo es venenoso cuando: a) el diámetro del sombrero es menor a 4.06 cms y su superficie es: correosa, escamosa, sedosa, fibrosa, arrugada, carnosa, surcos. b) si el diámetro del sombrero es mayor a 4.06 y la superficie del sombrero es arrugada. c) si el diámetro del sombrero es mayor a 4.06 cms y la superficie del sombrero es lisa, el color del tallo blanco, y el color del sombrero gris o marrón. d) si el diámetro del sombrero es mayor a 4.06, la superficie del sombrero es escamosa, el color del sombrero marrón, brillante, gris, verde, rosado, púrpura, rojo, amarillo, azul o negro.

Cuarto. la superficie del tallo es pegajosa, y además se cumple una de las siguientes condiciones: a) la altura del tallo es menor 5.2 cms y el color del tallo es rojo, púrpura, rosado, brillante, marrón, verde, azul, naranja o negro. Si el color del tallo es gris o amarillo el ancho del tallo del hongo deberá ser superior a 8.22 cms o

2.08 cms respectivamente. b) si la altura del tallo es mayor a 5.2 cms, el diámetro del sombrero es menor que 1.73 cms el color del sombrero marrón y su superficie brillante o pegajosa. c) si el tamaño del sombrero es mayor que 1.73 cms, el color del tallo marrón, brillante, gris, verde, rosado, púrpura, rojo, azul o negro. d) si el tamaño del sombrero es mayor que 1.73 cms, el color del tallo amarillo y la superficie del sombrero es correosa, escamosa, brillante, fibrosa, sedosa, arrugada o carnosa. En este último caso si la superficie del sombrero es pegajosa, y su diámetro es mayor a 1.87 cms el hongo también es venenoso. Si el diámetro es menor a este valor (1.87 cms), es venenoso si el color del sombrero es: marrón, brillante, gris, verde, rosado, púrpura, rojo, blanco, naranja o negro.

Respecto a los grupos de interés (clúster) según se muestra en las tablas 15 y 16 se identifican 5 grupos en los que se clasifican los hongos como venenosos con un total de 23.200 registros. Dado que son 27.181 registros de hongos venenosos, este sistema de clasificación presenta una efectividad del 85.33%, no obstante, aunque su efectividad sea menor, simplifica el proceso de reconocimiento de un hongo venenoso. Algunos autores (Dong et al., 2021) por medio de un sistema de redes neuronales, realizan una clasificación de hongos empleando imágenes con una efectividad del 99.76% cuando se entrena y valida con todo el conjunto de datos. Sin embargo, en dicho artículo, no se identifican variables similares a las empleadas en este artículo, tal vez por que la clasificación se basa en imágenes.

Otros autores como (Rahman et al., 2022), emplean un sistema de monitoreo inteligente y automatización agrícola en un sistema de aprendizaje con el fin de identificar los hongos venenosos de los que no lo son. Este modelo se basa en clasificadores como: arboles de decisión, búsqueda taboo, máquinas de soporte vectorial, clasificadores bayesianos, bosques de decisión, se logra así una clasificación del 100%, cuando se emplean arboles de decisión, búsqueda Taboo o bosques de decisión. En este proceso entre las variables más importantes se identifican: Color branquias, color de las esporas, población, tamaño de las branquias, raíz del tallo, hábitat, moretones, superficie del tallo, siendo esta última variable la única que ese y este trabajo identifican entre sus principales, esto quizás se deba a las metodologías empleadas, no obstante, este último aspecto, demuestra que existe más de un conjunto de variables que permiten lograr una clasificación óptima. Además, en este artículo investigativo también se logra una efectividad del 100% (Tabla 13) cuando se emplean las técnicas, IB1, IBK, bosques de decisión y arboles aleatorios, entrenando y validando con todo el sistema, sobre la base de datos inicial.

CONCLUSIONES

La metodología presentada en este artículo permite identificar si un hongo es venenoso o no con una efectividad de 99.76% y 7 variables: Diámetro sombrero, superficie sombrero, color sombrero, altura tallo, grosor tallo, superficie tallo, color tallo. No obstante, al ser comparado con otros estudios se encuentra que este trabajo presenta una efectividad de clasificación similar a los mismos. Es importante resaltar que en los estudios con los que se realizó una comparación, la mayoría de las variables independientes encontradas son diferentes, a las encontradas en este estudio, quizás debido a los procesos metodológicos empleados. Esto significa que existen diferentes conjuntos con los que se puede lograr una clasificación, con un porcentaje de éxito del 100%. Por último, se resalta que la metodología de clúster, aunque presenta buenos resultados de efectividad (85.33%), y simplifica el proceso de clasificación, no es tan efectiva en comparación con otras metodologías empleadas, aunque simplifica su proceso de clasificación.

AGRADECIMIENTOS

Se agradece a: Dua, D. y Graff, C., UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, (2019).

REFERENCIAS

- Ardabili, S., Mahmoudi, A., y otros dos autores, Modeling and Comparison of Fuzzy and On/Off Controller in a Mushroom Growing Hall, <http://dx.doi.org/10.1016/j.measurement.2016.04.050>, Measurement, 90, 127–134 (2016)
- Arunachalam, K., Puthanpura, S.S., y Yang, X., A Concise Review of Mushrooms Antiviral and Immunomodulatory Properties that May Combat Against COVID-19, <https://doi.org/10.1016/j.focha.2022.100023>, Food Chemistry Advances, 1, 100023 (2022)
- Braat, N., Koster, M., y Wösten, H., Beneficial Interactions Between Bacteria and Edible Mushrooms, <https://doi.org/10.1016/j.fbr.2021.12.001>, Fungal Biology Reviews, 39, 60 -72 (2022)
- Cano-Estrada, A., y Romero-Bautista, L., Valor Económico, Nutricional y Medicinal de Hongos Comestibles Silvestres, <http://dx.doi.org/10.4067/S0717-75182016000100011>, Rev Chil Nutr., 43(1), 75-80 (2016)
- Castrillón, O., Predicción del Divorcio por Técnicas de Minería de Datos, <http://dx.doi.org/10.4067/S0718-07642021000500111>, Información Tecnológica, 32(5), 111-120 (2021)

- Dong, J., Zhang, J., y otros dos autores, Deep Learning for Species Identification of Bolete Mushrooms with two-Dimensional Correlation Spectral (2DCOS) Images, <https://doi.org/10.1016/j.saa.2020.119211>, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 249, 119211 (2021)
- Dua, D., y Graff, C., UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science (2019)
- Lim, C., Chhabra, N., y otros cuatro autores, Atlas of Select Poisonous Plants and Mushrooms, <http://dx.doi.org/10.1016/j.disamonth.2015.12.002>, Disease-a-Month, 62, 41–66 (2016)
- Lua, C., y Liaw, J., A Novel Image Measurement Algorithm for Common Mushroom Caps Based on Convolutional Neural Network, <https://doi.org/10.1016/j.compag.2020.105336>, Computers and Electronics in Agriculture, 171, 105336 (2020)
- Moumita, S., y Das, B., Assessment of the Prebiotic Potential and Bioactive Components of Common Edible Mushrooms in India and Formulation of Synbiotic Microcapsules, <https://doi.org/10.1016/j.lwt.2021.113050>, LWT - Food Science and Technology, 156, 113050 (2022)
- Oliveira, V., Almeida, A., y otros siete autores, A New Circular Economy Approach for Integrated Production of Tomatoes and Mushrooms, <https://doi.org/10.1016/j.sjbs.2021.12.058>, Saudi Journal of Biological Sciences, 29, 2756–2765 (2022)
- Panda, S., y Luyten, W., Medicinal Mushrooms: Clinical Perspective and Challenges, <https://doi.org/10.1016/j.drudis.2021.11.017>, Drug Discovery Today, 27(2), 636-651 (2022)
- Peter, M., Liu, Z., y otros cinco autores, Computational Intelligence and Mathematical Modelling in Chanterelle Mushrooms' Drying Process Under Heat Pump Dryer, <https://doi.org/10.1016/j.biosystemseng.2021.10.002>, Biosystems Engineering, 212, 143 -159 (2021)
- Rahman, H., Faruq, O., y otros ocho autores, IoT Enabled Mushroom Farm Automation With Machine Learning to Classify Toxic Mushrooms in Bangladesh, <https://doi.org/10.1016/j.jafr.2021.100267>, Journal of Agriculture and Food Research, 7, 100267 (2022)
- Schunko, C., Li, X., y otros cinco autores, Local Communities' Perceptions of Wild Edible Plant and Mushroom Change: A Systematic Review, <https://doi.org/10.1016/j.gfs.2021.100601>, Global Food Security, 32, 100601 (2022)
- Saetang, N., Amornlerdpison, D., y otros tres autores, Processing of Split Gill Mushroom as a Biogenic Material for Functional Food Purpose, <https://doi.org/10.1016/j.bcab.2022.102314>, Biocatalysis and Agricultural Biotechnology, 41, 102314 (2022)
- Tian, R., Liang, Z., y otros dos autores, Analysis of Aromatic Components of two Edible Mushrooms, *Phlebopus portentosus* and *Cantharellus yunnanensis* Using HS-SPME/GC-MS, <https://doi.org/10.1016/j.rechem.2022.100282>, Results in Chemistry, 4, 100282 (2022)
- Valencia, M., Correa, J., y Díaz, F., Métodos Estadísticos Clásicos y Bayesianos para el Pronóstico de Demanda. Un Análisis Comparativo, <https://doi.org/10.15446/rev.fac.cienc.v4n1.49775>, Revista Facultad de Ciencias Universidad Nacional de Colombia, 4(1), 52 -67 (2015)
- Wagner, D., Heider, D., y Hattab, G., Mushroom Data Creation, Curation, and Simulation to Support Classification Tasks, <https://doi.org/10.1038/s41598-021-87602-3>, Sci Rep, 11, 8134 (2021)
- Wang, C., Wang, Y., y otros nueve autores, A Self-Floating and Integrated Bionic Mushroom for Highly Efficient Solar Steam Generation, <https://doi.org/10.1016/j.jcis.2021.12.064>, Journal of Colloid and Interface Science, 612, 88–96 (2022)
- Wen, X., y Jing, P., Dietary Cerebrosides in Seven Edible Mushrooms: One Step Detection, Quantification, and Si-SPE Assisted Isolation, <https://doi.org/10.1016/j.jfca.2022.104452>, Journal of Food Composition and Analysis, 108, 104452 (2022)
- Witten, I., Frank, E., y otros dos autores, Data Mining Practical Machine Learning Tools and Techniques, Morgan and Kaufman publication (Elsevier), ISBN-13: 978-0128042915, Cambridge, USA (2017)
- Xu, M., Zhu, S., y otros cuatro autores, Effect of Selenium on Mushroom Growth and Metabolism: A Review, <https://doi.org/10.1016/j.tifs.2021.10.018>, Trends in Food Science & Technology, 118, 328–340 (2021)
- Zhao, Z., Fan, T., y otros ocho autores, A Simple Derivatization Method for Simultaneous Determination of Four Amino Group-Containing Mushroom Toxins in Mushroom and Urine by UPLC-MS/MS, <https://doi.org/10.1016/j.foodcont.2021.108720>, Food Control, 137, 108720 (2022)
- Zotti, M., Zappatore, S., y Tosa, M., A Decision Support System for the Management of Accidental Mushroom and Plant Poisoning, [https://doi.org/10.1016/s0014-827x\(01\)01103-x](https://doi.org/10.1016/s0014-827x(01)01103-x), Il Farmaco, 56(5-7), 391-395 (2001)