

Primera entrega de proyecto

Gabriel Eduardo Camargo Garcia
Francisco Javier Silva Cadavid

Asignatura

Deep Learning

Docente

Raúl Ramos Pollán



Universidad de Antioquia

Facultad de Ingeniería

Medellín

2023

Contexto de aplicación

En la naturaleza, se encuentran 5 reinos taxonómicos de los seres vivos, los cuales son: reino Protista, reino Monera, reino Plantae, reino Animalia y reino Fungi. Este último reino es el cual pertenecen los hongos, los cuales son los grandes descomponedores de la naturaleza. Los hongos se encargan de reciclar la materia orgánica con la que interactúan. Como seres omnívoros, tenemos la capacidad de ingerir todo tipo de alimento, y dentro de nuestra dieta están los hongos, pero nos encontramos que la mayoría de estos son venenosos. Durante toda la evolución del ser humano, hemos diseñado técnicas para identificar aquellos que nos produzcan intoxicaciones graves, leves o nulas. Las graves que pueden matarnos, los leves que producen alucinaciones y las nulas que se usan en la gastronomía.

Objetivo de machine learning

Predecir si un hongo es comestible o venenoso en función de sus características físicas haciendo uso de las métricas y algoritmos comúnmente usados en problemas de clasificación.

Dataset

El dataset de clasificación de hongos es un dataset de clasificación binaria que se utiliza para identificar si un hongo es comestible o venenoso en función de sus características físicas. Este dataset consta de 8124 instancias, cada una con 23 atributos que describen las características físicas de un hongo. El tamaño del dataset es de 374 kB y se encuentra en formato CSV. La distribución de las clases en este dataset es casi equilibrada, con 4208 instancias para setas comestibles y 3916 instancias para setas venenosas.

Los atributos incluyen:

- Class: determina si el hongo es comestible o venenoso, con opciones de venenoso y comestible.
- Cap-shape: forma del píleo (sombrero) del hongo, con opciones de campana, cónico, convexo, plano, nudoso, hundido.
- Cap-surface: textura del píleo, con opciones de fibroso, surcado, escamoso, suave.
- Cap-color: color del píleo, con opciones de café, canela, gris, ante, verde, rosado, morado, rojo, blanco, amarillo.

- Bruises: ¿El hongo posee manchas?, con opciones de sí, no.
- Odor: tipo de olor del hongo, con opciones de almendra, anís, creosota, de pescado, acre, rancio, sin olor, mordaz, picante.
- Gill-attachment: morfología de las láminas del hongo, con opciones de adjunto, decurrente, libre, onduladas.
- Gill-spacing: distribución de las láminas, con opciones de cercano, pegado, distante.
- Gill-size: tamaño de las láminas, con opciones de ancho, estrecho.
- Gill-color: color de las láminas, con opciones de negro, café, ante, chocolate, gris, verde, naranja, rosado, morado, rojo, blanco, amarillo.
- Stalk-shape: forma del estípite del hongo, con opciones de ancho, estrecho.
- Stalk-root: raíz del estípite, con opciones de bulboso, palo, taza, igual, rizomorfo, arraigada, sin raíz.
- Stalk-surface-above-ring: textura del estípite encima del anillo, con opciones de fibroso, surcado, escamoso, suave.
- Stalk-surface-below-ring: textura del estípite debajo del anillo, con opciones de fibroso, surcado, escamoso, suave.
- Stalk-color-above-ring: color del estípite encima del anillo, con opciones de café, canela, gris, ante, verde, rosado, morado, rojo, blanco, amarillo.
- Stalk-color-below-ring: color del estípite debajo del anillo, con opciones de café, canela, gris, ante, verde, rosado, morado, rojo, blanco, amarillo.
- Veil-type: tipo del velo del hongo, con opciones de parcial, universal.
- Veil-color: color del velo, con opciones de café, naranja, blanco, amarillo.
- Ring-number: número de anillos, con opciones de cero, uno, dos.
- Ring-type: tipo de anillo, con opciones de telaraña, evanescente, enardecido, largo, colgante, revestido, separado.
- Spore-print-color: color de la esporada que libera el hongo, con opciones de negro, café, ante, chocolate, verde, naranja, morado, blanco, amarillo.
- Population: distribución de la población del hongo, con opciones de abundante, agrupada, numerosa, repartida, solitaria, poca.
- Hábitat: hábitat del hongo, con opciones de planta, hoja, prado, tierra

Métricas de desempeño

La métrica de desempeño principal será el área bajo la curva ROC, que calcula la calidad general del modelo en función de la tasa de verdaderos positivos y la tasa de falsos positivos. Debido a que es un problema de clasificación, es perfecta distinguir entre las clases. La precisión, la sensibilidad, la especificidad y el F1-score también serán útiles para evaluar este desempeño. Además, en términos de negocio, el costo de predecir un hongo venenoso como comestible es mayor que el costo de predecir un

hongo comestible como venenoso, ya que la ingestión de hongos venenosos puede ser peligrosa.

Referencias y resultados previos

Este dataset ha sido utilizado en muchos trabajos de investigación y proyectos de machine learning, y ha producido varios modelos precisos para predecir la comestibilidad de hongos. Algunos de los resultados previos incluyen el uso de random forest, máquinas de vectores de soporte, redes neuronales y Naive-Bayes. En general, se destaca el modelo basado en el algoritmo Naive-Bayes, como se visualiza en el siguiente enlace: <https://www.kaggle.com/code/upadorprofzs/bayesian-learning-basics-tutorial>

En términos de negocio, el costo de predecir un hongo venenoso como comestible es mayor que el costo de determinar un hongo comestible como venenoso. En el caso de los hongos venenosos, el costo de producirlos como comestibles podría resultar en envenenamiento y riesgo para la salud humana. Por lo tanto, minimizar la salida de falsos negativos (predecir un hongo venenoso como comestible) es imperativo para garantizar la seguridad del que lo va a consumir.

Bibliografía

- UCI Machine Learning. (n.d.). Mushroom Classification. Recuperado el 26 de marzo de 2023, de <https://www.kaggle.com/uciml/mushroom-classification>
- Upadhyay, A. (2021). Bayesian Learning Basics Tutorial. Kaggle. Recuperado el 26 de marzo de 2023, de <https://www.kaggle.com/code/upadorprofzs/bayesian-learning-basics-tutorial>