Mushroom Edibility

Data Science Boothcamp

Project 4 Group 10

Team Members

- Andrew Garza
- Sarah Ruth
- Kelsey Mersinas
- Laura Zapata

Purpose

The purpose of this project is to classify mushrooms as either edible or poisonous based on key characteristics like color, shape, and size. This will provide a practical tool for safer foraging, helping individuals quickly and accurately identify mushrooms in the wild. Additionally, we aim to gain insights into how these characteristics influence mushroom safety and edibility.

Inspiration

We were inspired by the critical need for safe mushroom identification and existing tools that address this challenge. In particular, we looked at examples in Tableau that demonstrated the potential for interactive presentations of mushroom data. Our project aims to enhance these concepts by incorporating machine learning and creating a user-friendly solution that improves the assessment of mushroom edibility.

Design Concepts

Our design concepts focus on creating an intuitive and engaging user experience for mushroom identification. Key elements include:

- **User-Friendly Interface:** We have designed a clean and straightforward layout that allows users to navigate easily and access information quickly.
- **Interactive Features:** Users can input mushroom characteristics to receive instant feedback on edibility, enhancing engagement and understanding.
- **Visual Hierarchy:** Important information is prioritized through size and color, ensuring users can quickly identify critical details.
- **Dynamic Visualizations:** We have implemented dynamic visualizations that effectively represent data trends, making it easier for users to grasp the relationships between mushroom features and their edibility.

Research Questions

1. What common features do poisonous mushrooms share?

- Poisonous mushrooms often exhibit specific patterns in characteristics like color and shape that indicate toxicity.

2. Does the season affect mushroom edibility?

- Yes, the season significantly impacts mushroom edibility. For instance, Fall shows a higher number of edible mushrooms compared to poisonous ones, but foragers must remain cautious as both types can be present.

3. Which matters more for determining edibility: color or size?

- Both color and size are important in identifying mushroom edibility, but ongoing analysis suggests that one may be more critical than the other in determining whether a mushroom is safe to eat.

Conclusion

Our project focuses on enhancing mushroom identification through machine learning. We explore key questions about the differences between edible and poisonous mushrooms. Our interactive Tableau visualizations help users understand these insights easily. By combining these findings with a user-friendly design, we aim to empower foragers to identify mushrooms confidently and safely, ultimately fostering a deeper appreciation for the diverse world of mushrooms. Machine learning is an excellent tool for educational and identification purposes that should be used in conjunction with human expertise.

Limitations/Bias

The data used to determine mushroom edibility was missing an important feature: spore color, which is crucial for identifying similar mushrooms but wasn't included, likely due to the time it takes to release spores. Adding spore and geographical data would improve the model's accuracy. While the Random Forest model performed well, subtle differences between species can lead to errors, and the model's complexity may be a challenge for systems with limited resources. Enhancing the model with more data would improve both its accuracy and usability.

Future Work

- Create a responsive web design for mobile use.
- Adding additional values to the data for the model to train on to increase accuracy and usability.
- Including geographical data to narrow down species
- Clean and engineer the original dataset and run the machine learning models again.

Works Cited

(n.d.). A banner of various mushrooms. Vector illustration of fungus. Drawing of voluminous fungi. Pro Vector. Vecteezy. Retrieved September 30, 2024, from https://www.vecteezy.com/vector-art/14315476-a-banner-of-various-mushrooms-vector-illustration-of-fungus-drawing-of-voluminous-fungi

(n.d.). Black mushroom icon. ICONSDB.COM. Retrieved September 30, 2024, from https://www.iconsdb.com/black-icons/mushroom-icon.html

R. (2024, February 15). What are the Advantages and Disadvantages of the Random Forest? Geeksforgeeks.org. Retrieved October 1, 2024, from https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/

Sawhney, P. (April 2024). Mushroom_cleaned_csv, Version 1. https://www.kagglecom/datasets/prishasawhney/mushroom-dataset/data
Sci Rep 11, 8134. Retrieved September 27, 2024, from https://doi.org/10.1038/s41598-021-87602-3

Shin, T. (2023, November 7). Understanding Feature Importance in Machine Learning. Builtin.com. Retrieved October 1, 2024, from https://builtin.com/data-science/feature-importance#:~:text=Feature%20 importance%20is%20a%20 step,to%20 predict%20a%20 certain%20variable.

Wagner, D., & Hattab, G.(2021).Secondary Mushroom [Dataset]. UCI Machine Learning Repository. Retrieved September 27, 2024, from https://doi.org/10.24432/C5FP5Q

Wagner, D., Heider, D. & Hattab, G. (2021).Mushroom data creation, curation, and simulation to support classification tasks. UCI Machine Learning Repository. Retrieved September 27, 2024, from https://www.nature.com/articles/s41598-021-87602-3