Project 4 - Group 10 - Proposal

Andrew Garza, Sarah Ruth, Kelsey Mersinas, Laura Zapata

Link to gitHub: https://github.com/KelseyMersinas/Project-4-Group-10

For Project 4, we selected a dataset on wild mushroom classification aimed at identifying edibility. This dataset, sourced from the UC Irvine Machine Learning Repository, contains over 50,000 entries representing various mushroom species. Given that there are over 14,000 species of mushrooms worldwide, ranging from perfectly edible to deadly, accurate mushroom identification is critical for foraging and safe consumption. Machine learning serves as an effective tool for quickly classifying mushrooms based on their characteristics, providing valuable information to individuals in the field.

Dataset : https://www.kaggle.com/datasets/prishasawhney/mushroom-dataset

The dataset consists of over 50,000 rows with a fairly balanced distribution of classes, exhibiting approximately a 55/45 percent split between poisonous and edible mushrooms. This balance will enhance the learning capabilities of our model and help mitigate bias during training. Key features in the dataset include: color, shape, gills, and size.

Using Tableau for visualization, we aim to explore several research questions related to mushroom edibility, including:

1. Characteristics of Poisonous Mushrooms: What common features are shared among poisonous mushroom species?

2. Seasonality and Edibility: Does the season in which a mushroom is found affect its edibility?

3. Impact of Color and Size: Which of the two characteristics—color or size—plays a larger role in determining the edibility of mushrooms?

We will create several interactive graphs and visualizations to illustrate our findings and facilitate deeper exploration of these inquiries.

Through our analysis, we anticipate gaining insights into the relationship between mushroom characteristics and edibility, which can contribute to safer foraging practices.

Additionally, our findings may enhance the understanding of how different environmental factors influence mushroom safety.

For roles we plan to all clean and scale the data, assemble our web app and slides and will break off individually for visualizations and specific pages on the web app.

Kelsey: ML model and training, flask initiation, works cited, about us, python anywhere

Ruth: ML model and training, embedded report web app page, home page

Laura: dashboard set up, 2-3 visualizations, tableau web app page

Andrew: 2nd dashboard set up, 2-3 visualizations, 2nd tableau web app page