Mushroom Classifier

Linn Vågenes Flesland, Henning Øinas

DESCRIBE THE PROBLEM

SCOPE

The goal of this project is to build a machine-learning model that classifies mushrooms as edible or poisonous based on physical characteristics. The purpose is to demenstrate the end-to-end ML workflow, including data gathering, preprocessing, modeling, and deployment.

METRICS

Model performance is measured using:

- Accuracy and F1-score (classification metrics).
- The Random Forest achieved x% accuracy.

In a buisiness context, the key metric would be to minimize false negatives (poisonous predicted as edible). Overall accuracy is sufficient to indicate success.

DATA

The dataset is the Mushroom dataset (ID 73) from the UCI Machine Learning Repository, accessed using the Python package `ucimlrepo`.

- **Type**: Tabular, 8124 rows, 22 catagorical features.
- **Label**: “class” - edible (`e`) or poisonous (`p`).
- **Missing values**: One feature (`stalk-root`) uses `?` to mark unknown values. These are replaced with the catagory “unknown”.
- **Privacy**: No personal or sensitive data.

MODELING

- **Approach**: Supervised classification.
- **Model**: RandomForestClassifier (100 trees, fixed random seed).
- **Baseline**: Accuracy from a simple majority-class prediction (~50%).
- **Final model**: Random Forest with one-hot-encoded inputs → ~100% test accuracy.

Why we chose Random Forest:

- Handles purely categorical features well after encoding.
- Is robust to noise and overfitting
- Provides built-in feature importance metrics

Evaluation uses an 80/20 train-test split.

DEPLOYMENT

Deployment is handled using Streamlit Cloud, allowing users to run the app directly in a web browser as well as run it locally.

REFERENCES

- UCI Machine Learning Repository - Mushroom Dataset: https://archive.ics.uci.edu/dataset/73/mushroom
- scikit-learn documentation - RandomForestClassifier & Pipeline

- Streamlit documentation
- Streamlit as deployment
- Philip Truong ucimlrepo package for dataset access, available on PyPi and GitHub
- OpenAI ChatGPT - guidance and coding assistance