Good to eat or deadly poison?
Machine Learning Techniques for Mushroom Classification

Abstract

Applying a variety of supervised machine learning models and approaches, this project investigates the feature importance and model performance on classification of edible and poisonous mushrooms based on a diverse set of ecological and physical attributes. We aim to evaluate how well different modeling strategies can distinguish poisonousness and identify which features are most contributory in accurate predictions. Study results show that Ensemble method achieves the best classification performance with best accuracy and F2 score, while attributes such as odor, gill size and stalk surface appear to be the most influential indicator of poisonous mushrooms.

Key Words: mushroom classification; machine learning; supervised learning; ensemble methods; feature importance

I. Introduction

With thousands of varieties of mushrooms differing in appearance, ecological traits and habitats, distinguishing between edible and poisonous mushrooms becomes a difficult and complex task, even for experienced foragers. As a result, we found that accurately predicting edible varieties from poisonous ones given observed features is a critical and meaningful task. In our case, misclassification can have severe health consequences, so beyond the biological meaning, this problem also highlights how to effectively apply machine learning classification models under extremely low misclassification tolerance.

Mushroom identification is inherently pattern-based, yet these patterns are often non-linear and unintuitive to humans, machine learning offers a systematic way to uncover these relationships and implications from the combination of feature values. We selected both generative and discriminative approaches, as well as tree-based ensemble methods to capture the non-linearity. For each model, we evaluate their performance using several metrics and compare for the best one. At the end, we performed a stack ensemble to combine each models' predictions. The dataset we selected contains only categorical variables, each corresponding to an environmental or morphological characteristic of an observed mushroom, such as cap attributes, odor, gill configuration, and a target variable indicating the poisonousness of this mushroom, more details about the dataset would be discussed during the next section.

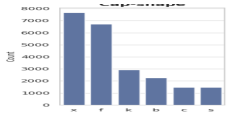
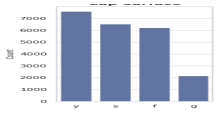
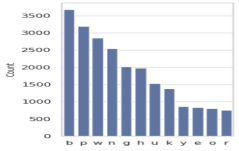
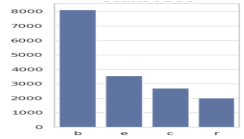
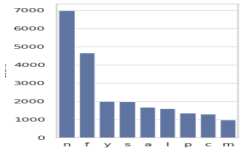
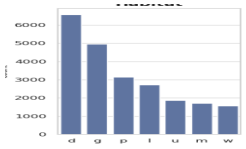
We aim to determine the features that contribute most strongly for distinguishing poisonous or edible mushrooms and identify which model delivers the best predictive performance. Our findings offer practical insights into designing reliable classification systems for biological data and tasks.

II. Data Overview and EDA

This is a binary classification problem. The goal is to build a model that predicts whether a mushroom is edible(e) or poisonous(p). The dataset contains more than 25000 rows, this is a large, enhanced dataset.

Our Target variable is '*class*': a categorical type. And it's well-balanced, containing a high number of both edible and poisonous samples, with slightly more edible ones.

The table below give us a preview of some of the outstanding features:

Features	**Description**	**Distribution**
cap-shape	Shape of the cap (bell, conical, convex, flat…)	
cap-surface	Surface texture of the Cap (fibrous, grooves, scaly, smooth)	
gill-color	Color of the gills (black, brown, buff, chocolate, gray, green, orange, pink, purple, red, white, yellow)	
stalk-root	Root of the stalk (bulbous, club, equal, rhizomorphs, rooted, missing)	
spore-print-color	Color of the spore print (black, brown, buff, chocolate, green, orange, purple, white, yellow)	
odor	Odor (almond, anise, creosote, fishy, foul, musty, none, pungent, spicy)	
habitat	Habitat where the mushroom was found (grasses, leaves, meadows, paths, urban, waste, woods)	

Most features are distributed evenly regarding whether they are edible or poisonous. The complete set of features we used is in **Table 1** in Appendix.

By checking the Cramer's V correlation heatmap (**Figure 1**, Appendix) , which measures the strength of association between categorical features. *'odor'* is the best predictor and has the strongest association with *'class'*, showing a value of 0.55, which is significantly higher than any other features. Other features are weak predictors, such as *'spore-print-color'*, *'gill-size'*, and *'stalk-surface-above-ring'*, which have very weak association with target. The rest of the map indicates that most features are not associated with each other, which is excellent for later models because they all provide unique, non-redundant information. The only feature we removed during data processing is *'veil-type'* as it contains only one unique class across the entire dataset with zero variability.

III. Modeling Results

We implemented a wide range of supervised learning models, we began with LDA and QDA, two classical generative models, and then applied Logistic Regression, KNN and SVM with RBF kernels. In addition to these standalone methods, we also explored tree-based ensemble models. Random Forest aggregates many bootstrapped trees for variance reduction and Gradient Boosting sequentially builds shallow trees to correct previous prediction errors. Finally, we constructed a Stack Ensemble model, combining multiple base learners to integrate their predicted probabilities for a meta Logistic Regression learner to gain better performance.

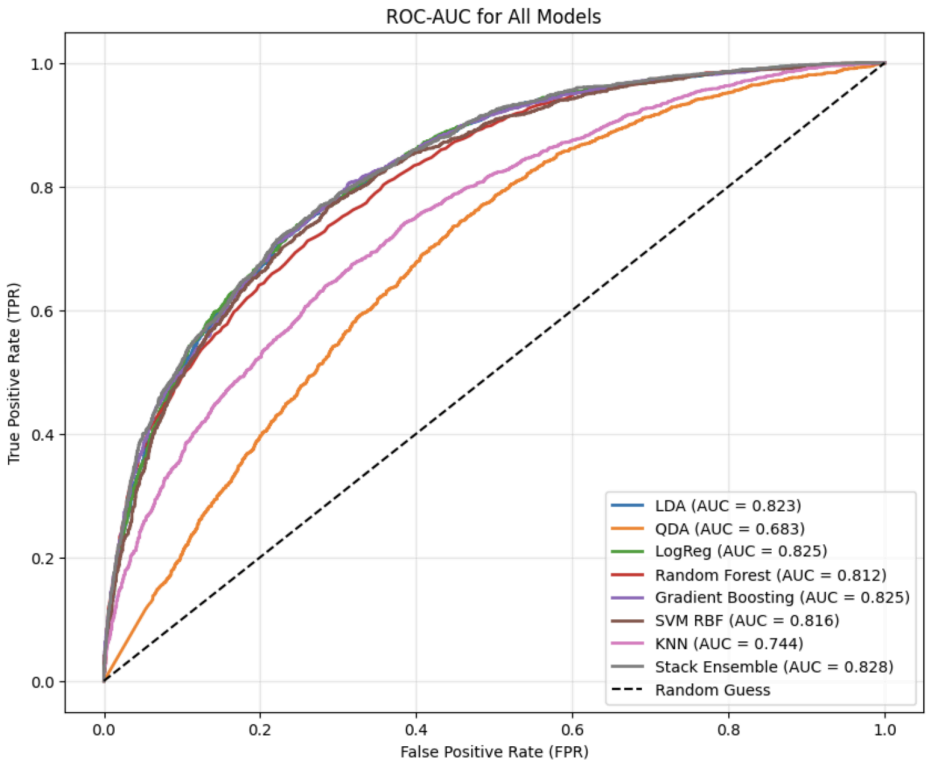


Figure 2

We then evaluated each model using several performance metrics, including accuracy and ROC-AUC analysis, because the task involves real health and safety risks, the consequences

of prediction errors are asymmetric: a False Negative (misclassifying a poisonous mushroom as edible) is far more harmful than a False Positive. Therefore, our particular emphasis is placed on maintaining a low FNR and strong F2 score. In addition to training each classifier, we also tuned the probability decision threshold for every model to achieve balance between FNR, F2 score and overall accuracy. The final model performance summary is in **Table 2** in Appendix, ROC-AUC analysis results are presented in **Figure 2**.

From the model performance metrics and AUC scores, most models exhibit broadly similar behaviour, with little dramatic deviations in overall predictive performance. LDA, Gradient Boosting, Logistic Regression and Stack Ensemble achieve almost consistently satisfactory performance. Stack Ensemble delivers the best predictive power but just marginally better than or equally good as the other models, suggesting that the underlying patterns are already well captured by those standalone models. Such consistency indicates that the classification task is relatively stable and we may have a handful of models selected for yielding reliable predictions. We would choose Stack Ensemble as our final model.

IV. Discussion of Model Implications

4.1 Best Model Evaluation

Final Model	F2	Accuracy	FPR	FNR
Stack Ensemble	0.82	0.70	0.45	0.11

Our final model has a F2 score of 0.82, which means our model is reliably catching most of the poisonous mushrooms. Even though it produces more false alarms as shown by FPR, it reflects a good trade-off between safety and usability. We also managed to maintain good accuracy through this balanced trainingset.

4.2 Feature Importance

In order to extract the feature importance from the Stacked Ensemble model, we calculate the Permutation Importance, which shuffles each feature at a time and measures the drop in performance.

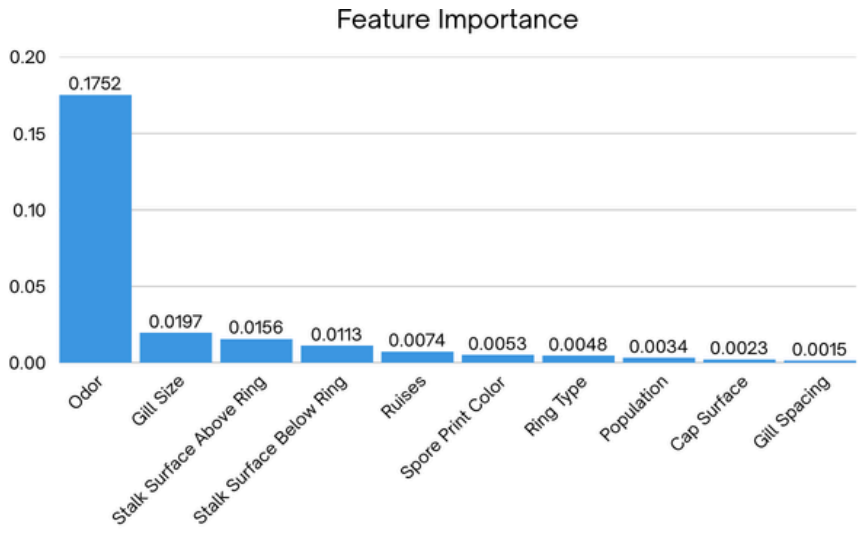


Figure 3

As shown in **Figure 3**, *"Odor"* is the most predictive feature followed by *"Gill Size"*, *"Stalk Surface Above Ring"*, and *"Stalk Surface Below Ring"*. As shown in **Figure 4**, the variation between the top 10 features are low between permutations, showing promising confidence in our feature ranking.

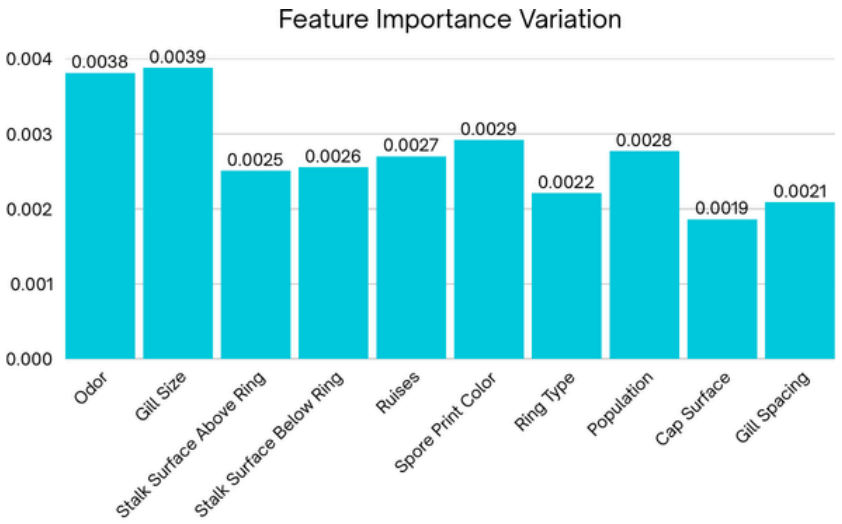


Figure 4

4.3 Discussion of Future Improvements

The dataset consists entirely of categorical variables, which limits the depth of information available from each feature. To strengthen the dataset, incorporating numerical features such as mushroom size, habitat temperature, or average life span could provide additional predictive signals.

Another improvement could be to consult domain experts to guide more meaningful encodings, although we currently rely on simple one-hot encoding, expert insight could enable the creation of risk-based numerical scores (e.g., assigning higher risk values to bright or unusual colors). However, we were unable to find reliable domain resources to support such numerical encodings in this project.

V. Conclusion

Therefore, the majority of the machine learning models we applied to classify mushrooms perform well on this categorical dataset. To be more specific, Stack Ensemble emerged as the best overall model, offering strong accuracy and a high F2 score while keeping the false-negative rate low—an essential requirement for safety-critical tasks. On the other hand, our feature analysis confirmed that odor is the strongest predictor, supported by structural features such as gill size and stalk surface attributes. These findings highlight that a few key characteristics contain most of the classification signal.

Overall, this study demonstrates the potential of machine learning in biological identification tasks and provides a strong foundation for building more robust and practical classification systems. In this case, we believe the model can help people avoid poisonous mushrooms by offering a quicker, and more data-driven safety check, improving decision-making for foragers and outdoor users.

Reference

A. Çeliktaş, "Classification with Machine Learning (Python) 'Mushroom Dataset'," *Become Better*,Feb.16,2024.[Online].Available: https://medium.com/becoming-for-better/classification-with-machine-learning-python-mushroom-dataset-790a275610df. [Accessed: Dec. 06, 2025]

C. Arnold, "Should I Really Eat That Mushroom?" *Towards Data Science*, Aug. 17, 2023. [Online].Available: https://towardsdatascience.com/should-i-really-eat-that-mushroom-9edeaa69d934/. [Accessed: Dec. 06, 2025]

N. J. Pinky, S. M. M. Islam, and R. S. Alice, "Edibility detection of mushroom using ensemble methods," *Int. J. Image Graph. Signal Process.*, vol. 11, no. 4, pp. 55-62, Apr. 2019, doi: 10.5815/ijigsp.2019.04.05.

Appendix

Figure 1: Cramer's V correlation heatmap

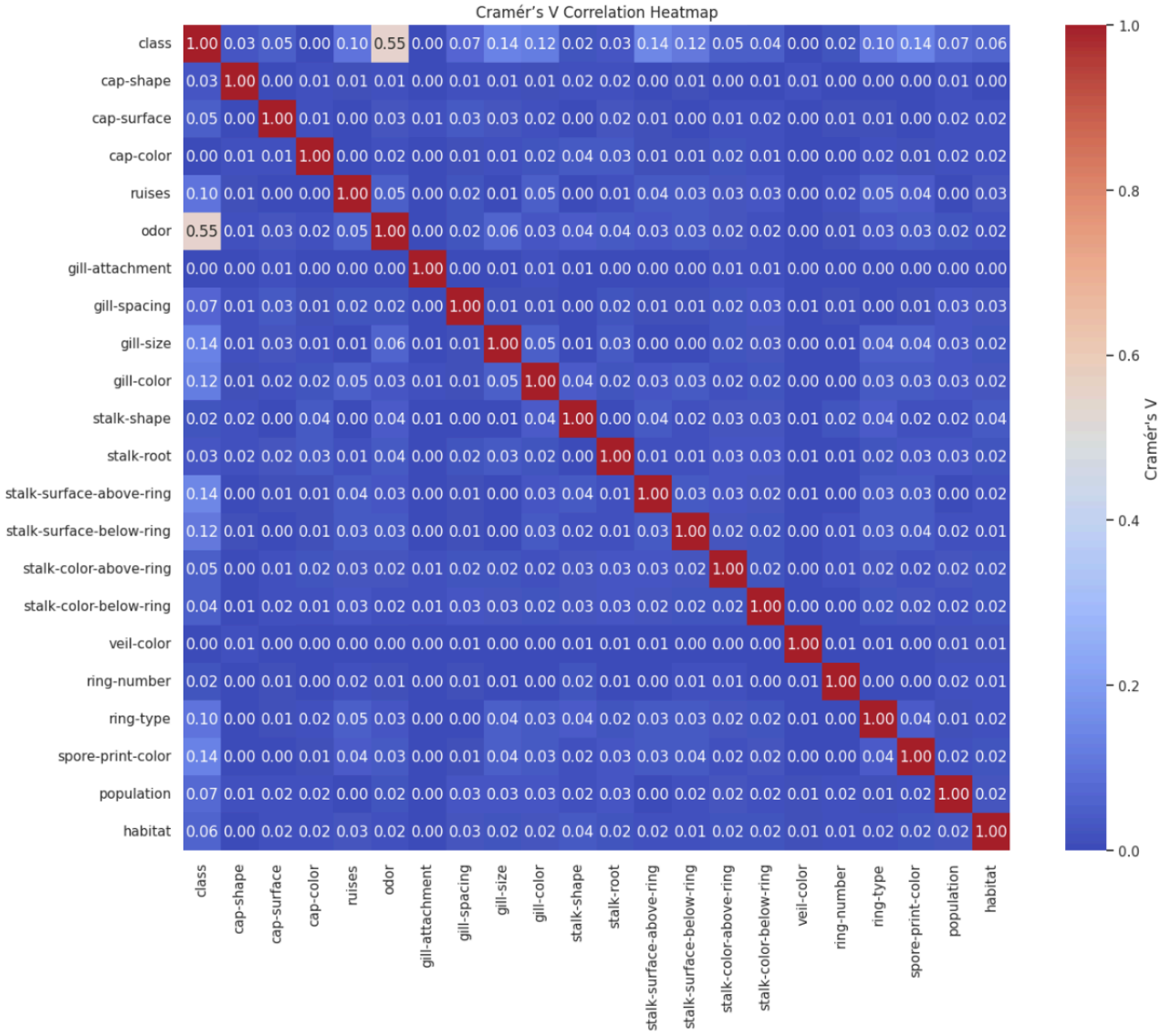


Table 1: Complete summary of all features

Color features	
cap-color	*Color of the cap* (brown, buff, cinnamon, gray, green…)
gill-color	*Color of the veil*(brown, orange, white, yellow)
stalk-color-above-ring	*Surface of the stalk above the ring* (brown, buff, cinnamon, gray, orange, pink, red, white, yellow)
stalk-color-below-ring	*Same as above* (brown, buff, cinnamon, gray, orange, pink, red, white, yellow)
veil-color	*Color of the Veil* (brown, orange, white, yellow)

spore-print-color	*Color of the spore print* (black, brown, buff, chocolate, green, orange, purple, white, yellow)
Shape & Structural Features	
cap-shape	*Shape of the mushroom cap* (bell, conical, convex, flat, knobbed, sunken)
gill-attachment	*Attachment of gills to the stalk* (attached, descending, free, notched)
gill-spacing	*Spacing between the gills*(close, crowded, distant)
gill-size	*Size of the gills* (broad, narrow)
stalk-shape	*Shape of the stalk*
stalk-root	*Root of the stalk* (bulbous, club, cup, equal, rhizomorphs, rooted, missing)
veil-type	*Type of Veil* (partial, universal)
ring-type	*Type of ring* (cobwebby, evanescent, flaring, large, none, pendant, sheathing, zone)
Texture&Surface Features	
cap-surface	*Surface texture of the cap* (fibrous, grooves, scaly, smooth)
stalk-surface-above-ring	*Surface of the stalk above the ring* (fibrous, scaly, silky, smooth)
stalk-surface-below-ring	*Surface of the stalk below the ring* (fibrous, scaly, silky, smooth)

Other/Contextual Attributes	
bruises	*Presence of bruises* (bruises, no)
odor	*Odor* (almond, anise, creosote, fishy, foul, musty, none, pungent, spicy)
ring-number	*Number of rings on the stalk* (none, one, two)
population	*Population density* (abundant, clustered, numerous, scattered, several, solitary)
habitat	*Habitat where the mushroom was found* (grasses, leaves, meadows, paths, urban, waste, woods)

Table 2: All model performance summary

Model	F2	Accuracy	TPR	FNR	FPR	TNR
LDA	0.80	0.71	0.88	0.12	0.43	0.57
QDA	0.77	0.61	0.86	0.14	0.60	0.40
Random Forest	0.81	0.68	0.90	0.10	0.50	0.51
Boosting	0.82	0.70	0.89	0.10	0.46	0.54
Logit Reg.	0.82	0.71	0.89	0.12	0.44	0.56
SVM	0.76	0.73	0.79	0.21	0.32	0.68
KNN	0.79	0.60	0.89	0.11	0.64	0.36
Stack Ensemble	0.82	0.71	0.89	0.11	0.45	0.55

Link of codes:

https://drive.google.com/drive/folders/1j7j36PRnLM8pPR9Hr2ubDWbaWouPpuW3?usp=sharing

Source of data:

https://www.kaggle.com/datasets/sakurapuare/mushroom-classification-enhanced

YouTube link of presentation:

https://youtu.be/-vT7vMlXtUc