International Journal of Scientific Research in Engineering and Management (IJSREM)
Conference - Volume 10 NCET -2026 SJIF Rating: 8.659 ISSN: 2582-3930

Mushroom Classification Using Machine Learning.

1st D.Lavanya
Dept of information Technology Syed Ammal Arts and Science College
Ramanathapuram, India
Praveenlava89@gmail.com

2nd B.Jeya Lakshmi
Dept of information Technology Syed Ammal Arts and Science College
Ramanathapuram, India
jeyamca14@gmailcom

3rd R.Rahini
Dept of information Technology Syed Ammal Arts and Science College
Ramanathapuram, India
Kamaleshtn65@gmail.com

Abstract— This paper presents a comprehensive analysis of mushroom classification using machine learning techniques. With over 45,000 known mushroom species worldwide, only a small fraction has been identified as edible. Many species are extremely toxic and pose significant health risks if consumed. This makes accurate identification challenging for non-experts. We apply supervised machine learning algorithms—including Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM)—to classify mushrooms as edible or poisonous based on morphological features. Experimental results demonstrate that SVM achieves superior classification accuracy compared to other algorithms, making it a reliable approach for automated mushroom identification.

Keywords: Mushroom classification, Support Vector Machine, Naïve Bayes, Logistic Regression, Machine Learning, Edible Detection

I. INTRODUCTION

Mushrooms are widely used as ingredients in the food industry and offer significant nutritional benefits. They contain essential nutrients such as calcium, phosphorus, and various vitamins, making them valuable components of balanced diet programs. Additionally, mushrooms have been studied for their potential medicinal properties, including cancer treatment support, viral eradication, immune system enhancement, and weight management applications.

Accurate mushroom identification is critical because many species are poisonous and can cause severe health complications if consumed. The morphological characteristics of mushrooms—such as cap shape, cap color, gill color, ring structure, and volva presence—serve as key discriminative features for classification. The cap is the top structure of the mushroom (often resembling a small umbrella), commonly appearing in brown, white, or yellow colors. The stem provides vertical support, while the volva represents the protective veil remaining after the mushroom emerges from the ground. Gills are the rib-like structures located

beneath the cap, and the ring (annulus) is the remnant of the partial veil after gill expansion. This study leverages these distinctive features to develop automated classification models using machine learning algorithms.

II. DATASET AND METHODOLOGY

The dataset utilized in this study comprises 8,124 mushroom instances with 23 categorical features describing morphological attributes. Key features include cap shape, cap surface, cap color, gill attachment, gill spacing, gill size, gill color, stalk shape, stalk surface, stalk color, veil type, veil color, ring number, ring type, spore print color, population, and habitat. The dataset was obtained from Kaggle's Mushroom Classification repository. All features are categorical and require appropriate encoding for machine learning model compatibility.

A. Data Preprocessing

The raw dataset undergoes systematic preprocessing to ensure model compatibility. Categorical variables are encoded using label encoding techniques to convert textual attributes into numerical representations. The dataset is partitioned into training and testing subsets using an 80:20 split ratio to facilitate unbiased model evaluation. Feature scaling is applied where necessary to normalize the range of independent variables.

B. Python Libraries and Tools

The implementation utilizes the following Python libraries and packages: (1) NumPy for numerical computations, (2) Pandas for data manipulation and analysis, (3) Seaborn for statistical data visualization, (4) Matplotlib for graphical representation, and (5) Scikit-learn for machine learning algorithm implementation and model evaluation.

III. CLASSIFICATION ALGORITHM

This section describes the three supervised learning algorithms employed for mushroom classification.

A. Logistic Regression

Logistic Regression is a widely adopted supervised learning algorithm for binary classification tasks. It predicts the probability of a categorical dependent variable using a logistic function, outputting values between 0 and 1. The algorithm supports three variants: (1) Binomial—for binary outcomes (e.g., edible/poisonous), (2) Multinomial—for three or more unordered categories, and (3) Ordinal—for ordered categorical variables. For this study, binomial logistic regression is applied to classify mushrooms into edible or poisonous categories.

B. Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' theorem with strong independence assumptions between features. Despite its simplifying assumptions, it often performs well in high-dimensional datasets. The algorithm computes two primary probabilities: (1) Posterior probability—the probability of a class given the observed features, and (2) Likelihood probability—the probability of observing the features given a specific class. The classification process involves importing the dataset, preprocessing categorical variables, splitting data into training and testing sets, and applying the Scikit-learn implementation of the Naïve Bayes classifier.

C. Support Vector Machine (SVM)

Support Vector Machine is a robust supervised learning algorithm effective for both linear and non-linear classification. The algorithm identifies an optimal hyperplane that maximizes the margin between data points of different classes. Key concepts include: (1) Hyperplane—the decision boundary separating classes, (2) Support Vectors—data points closest to the hyperplane that influence its position, and (3) Margin—the distance between the hyperplane and nearest support vectors. SVM solves a convex optimization problem to find coefficients that maximize the margin while correctly classifying training instances. A larger margin typically correlates with lower generalization error, making SVM particularly effective for classification tasks with clear margin separation.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of each classification algorithm is evaluated using standard metrics including accuracy, precision, recall, and F1-score. The following table presents the classification accuracy achieved by each algorithm:

Table 1: Classification Accuracy Comparison

Algorithm	Accuracy (%)
Naïve Bayes	92.18
Logistic Regression	94.65
Support Vector Machine	100.00

The experimental results demonstrate significant performance variations across the three algorithms. Naïve Bayes achieved 92.18% accuracy, while Logistic Regression improved upon this with 94.65% accuracy. Support Vector Machine exhibited the highest performance at 100.00% accuracy, indicating perfect separation of edible and poisonous mushroom classes in the feature space. The superior performance of SVM can be attributed to its ability to maximize the margin between classes and handle high-dimensional categorical data effectively. Additionally, SVM demonstrated faster computational processing compared to the other algorithms, making it efficient for real-time classification applications.

The comparison reveals that SVM outperforms both Naïve Bayes and Logistic Regression in terms of classification accuracy. This suggests that the mushroom dataset features exhibit linear or near-linear separability, which SVM exploits optimally through its margin-maximization objective.

V. CONCLUSION

This study evaluated three machine learning algorithms for automated mushroom classification. Experimental results indicate that Support Vector Machine achieves the highest accuracy at 100.00%, outperforming both Logistic Regression (94.65%) and Naïve Bayes (92.18%). Furthermore, SVM demonstrated superior computational efficiency, processing classification tasks faster than comparative algorithms. These findings suggest that SVM is a reliable and efficient approach for automated edible/poisonous mushroom identification, potentially assisting non-experts in making safe consumption decisions. Future work may explore ensemble methods, deep learning architectures, and cross-dataset validation to enhance generalization capabilities.

V. REFERENCES

[1] Nagulwar, M., D. More, and L. Mandhare, "Nutritional properties and value addition ofmushroom: a review," The Pharma Innovation Journal, vol. 9, no. 10, pp. 395-398, 2020.

[2] Boa, E.R., Wild Edible Fungi: A Global Overview of Their Use and Importance toPeople. Rome: Food and Agriculture Organization of the United Nations, 2004.

[3] Kousalya, K., et al., "Edible mushroom identification using machine learning," in Proc.2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-6.

[4] Ismail, S., A.R. Zainal, and A. Mustapha, "Behavioural features for mushroomclassification," in Proc. 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2018, pp. 1-6.

[5] Mahesh, B., "Machine learning algorithms—a review," International Journal of Scienceand Research (IJSR), vol. 9, pp. 381-386, 2020.

[6] Ottom, M.A., N.A. Alawad, and K. Nahar, "Classification of mushroom fungi usingmachine learning techniques," International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 5, pp. 2378-2385, 2019.

[7] Hamonangan, R., M.B. Saputro, and C.B.S.D.K. Atmaja, "Accuracy of classificationpoisonous or edible of mushroom using naïve bayes and k-nearest neighbors," Journal of Soft Computing Exploration, vol. 2, no. 1, pp. 53-60, 2021.

[8] Halili, F. and F. Kamberi, "Performance analysis of classification algorithms: a casestudy of Naïve Bayes and J48 in Big Data," Applied Mathematics and Computation, vol. 2, no. 2, pp. 50-57, 2021.