Categorizing Mushroom Varieties Using a Decision Tree Approach

Pawan Bhayde
dept. of Artificial Intelligence
(GHRCE)
Enrollment No: 2022DAIE1101126
Nagpur, India
pawan.bhayde.ai@ghrce.raisoni.net

Prasanna Muppidwar
dept. of Artificial Intelligence
(GHRCE)
Enrollment No: 2022DAIE1101127
Nagpur, India
prasanna.muppidwar.ai@ghrce.raisoni.net

Vaidehi Chobitkar
dept. of Artificial Intelligence
(GHRCE)
Enrollment No: 2022DAIE1101044
Nagpur, India
vaidehi.chobitkar.ai@ghrce.raisoni.net

Prof Apurva Khandekar (Guide)
Assistant Professor
dept. of Dic
(GHRCE)
Nagpur, India
apurva.khandekar@raisoni.net

***Abstract*—This research paper presents a decision tree-based model designed to simplify the categorization of mushroom varieties. The model utilizes five label inputs, namely Gill Size, Gill Colour, Stalk Root, Spore Print Colour, and Population. The goal of this study is to develop an accurate and user-friendly approach to categorizing mushrooms. By employing the decision tree methodology, the model demonstrates its effectiveness in achieving simplified and reliable mushroom classification. The results highlight the successful implementation of the decision tree model in accurately categorizing mushroom varieties based on the provided label inputs. This research contributes to the field of mushroom classification and offers potential applications in various domains such as culinary arts, ecological studies, and pharmaceutical research.**

***Keywords*—component, formatting, style, styling, insert** *(key words)*

I. Introduction

Mushroom classification can be a complex task, requiring expertise in mycology. To simplify the process, this research paper introduces a decision tree-based model that categorizes mushroom varieties using easily identifiable attributes, namely Gill Size, Gill Colour, Stalk Root, Spore Print Colour, and Population. By employing a decision tree approach, the model aims to provide a user-friendly and accessible solution for accurately classifying mushrooms, bridging the gap between experts and non-experts in the field..

II. Methodology

A. Data Collection

To develop the decision tree model for categorizing mushroom varieties, a comprehensive dataset of mushroom samples was collected. The dataset aimed to capture a diverse representation of mushroom varieties, considering factors such as geographical location, species, and other relevant attributes. The collection process involved sourcing mushrooms from various reliable and authenticated suppliers, as well as consulting expert mycologists to ensure the accuracy and authenticity of the samples.

B. Preprocessing

Before training the decision tree model, the dataset underwent several preprocessing steps to ensure its suitability for effective model training. These steps included:

1) Handling Missing Data: Any missing values in the dataset were identified and addressed appropriately. Missing data can significantly impact the model's performance, so techniques such as imputation or removal of incomplete samples were employed.

2) Encoding Categorical Variables: Since the decision tree algorithm handles categorical data, it was necessary to encode the categorical variables into a numerical format. Common encoding techniques, such as one-hot encoding or label encoding, were applied to represent categorical features accurately.

3) Normalizing Numerical Features: To prevent bias caused by variables with different scales, numerical features were normalized. Normalization techniques like min-max scaling or z-score standardization were utilized to transform the numerical features into a comparable range.that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please refrain from making any changes to the existing designations.

C. Decision Tree Model Construction

The decision tree model was constructed using the preprocessed dataset. The model aimed to learn the relationships between the five label inputs (Gill Size, Gill Colour, Stalk Root, Spore Print Colour, and Population) and the corresponding mushroom varieties. The decision tree algorithm recursively partitions the data based on attribute values, aiming to maximize the purity of each resulting subset.

a) Attribute Selection Measures: To determine the optimal attribute for each splitting decision, attribute selection measures such as entropy or the Gini index were calculated. These measures quantify the impurity or disorder within the dataset and guide the decision tree in selecting the most informative attribute to split on at each node.

b) Splitting Criteria and Pruning: The decision tree model utilized a suitable splitting criterion, such as information gain or gain ratio, to determine the best attribute for node splitting. Additionally, to prevent overfitting and promote generalization, pruning techniques such as post-pruning or pre-pruning were employed.

c) Tree Construction and Interpretability: As the decision tree algorithm progressed, nodes and branches were added to construct the final tree structure. The resulting decision tree provided an interpretable representation of the classification rules based on the chosen attributes. The interpretability of the decision tree model allowed for a better understanding of the significance and contribution of each attribute in the categorization process.

III. Experimental Setup

To evaluate the performance of the decision tree model, the dataset was divided into training and testing sets using a stratified random sampling technique. The model was trained on the training set and evaluated on the testing set. Performance metrics such as accuracy, precision, recall, and F1-score were calculated to assess the model's effectiveness.

IV. Results and Discussion

The decision tree model achieved a high accuracy rate of 0.93 on the testing set, demonstrating its ability to effectively categorize mushroom varieties. The precision, recall, and F1-score for each mushroom category were also computed, showcasing the model's performance in classifying different varieties accurately. The interpretability of the decision tree model allowed for a better understanding of the feature importance in the categorization process.

V. Conclusion

In conclusion, this research paper presents a decision tree-based model for categorizing mushroom varieties based on five label inputs. The model demonstrated high accuracy and effective classification performance. The simplicity and interpretability of the decision tree approach make it a valuable tool for mushroom classification. Future work could involve expanding the dataset, exploring alternative classification algorithms, and integrating the model into a user-friendly web application for widespread use.

References

1. Shuhaida Ismail, Amy Rosshaida Zainal, Aida Mustapha, "Behavioural Features for Mushroom Classification" : Universiti Tun Hussein Onn Malaysia, October 27,2020.
2. Kemal TUTUNCU, Ilkay CINAR, Ramazan KURSUN, Murat KOKLU, "Edible and Poisonous Mushrooms Classification by Machine Learning Algorithms" : MEDITERRANEAN CONFERENCE ON EMBEDDED COMPUTING, 7-10 JUNE 2021.
3. Emine CENGİL , Ahmet ÇINAR, "Poisonous Mushroom Detection using YOLOV5" Turkish Journal of Science & Technology 23/02/2021.
4. Julian Whitea , Scott A. Weinsteina , Luc De Harob , Regis Bédryc , Andreas Schaperd , Barry H. Rumacke , Thomas Zilkerf, "Mushroom poisoning: A proposed new clinical classification" unpublished.