

Mushroom Classification using Machine Learning Algorithms: An Experimental Approach

R Thabassum Kowsar

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract— Mushrooms are a diverse group of fungi found in various ecosystems around the world. While many mushrooms are edible and safe for consumption, some can be highly toxic and pose a significant health risk if ingested. This research paper explores the development of machine learning models to accurately classify mushrooms as either edible or poisonous based on their characteristics. The dataset used in this study consists of 8124 instances of mushrooms, each described by 23 features such as cap shape, cap surface, cap color, bruises, odor, and more. Among the instances, 4208 are edible, and 3916 are poisonous. In this paper, we apply different machine learning classifiers, including Artificial Neural Networks and Logistic Regression, to build models that can effectively identify the edibility of mushrooms. We evaluate these models based on accuracy, precision, and recall to determine their performance in mushroom classification.

I. INTRODUCTION

The objective of mix learning is to help a model that isolates the information into the various classes, completely reason on referring to new models later on. Get-together learning systems rather produce various models. Given another model, the affiliation passes it to the amount of its different base models, gets their questions, and eventually later obliges them in some fitting way (e.g., averaging or projecting an investigating structure). Most of outfit learning systems is standard, material across wide classes of model sorts and learning attempts. Affiliation learning is a reasonable design that has progressively been embraced to join different learning assessments to other than engage figure exactness, overall [1][2]. Maybe the most astonishing spaces of evaluation in oversaw recreated knowledge have been to look at strategies for making exceptional outfits of understudies. The key disclosure is that outfits are a basic piece of the time incredibly more particular than the solitary understudies [3][4]. While organizing an affiliation learning system, correspondingly as picking the structure by which to achieve grouping in the base models and picking the joining procedure, one vital for pick the kind of base model and base model learning assessment to utilize. The joining structure could limit such base models that can be utilized.

Information tunneling arranges looking for beguiling models or information from giant information. It changes an enormous get-together of information into information. Information mining is a basic improvement during the time spent information transparency. The information mining has transformed into an interesting mechanical party seeing assessing information as demonstrated by substitute point of view and changing over it into gigantic and major data [6]. Information digging has been overall applied in the space of clinical finding, Interference ID structure, Preparing, Banking, Coercion divulgence. Get-together is a coordinated learning. Measure and graph in information mining are two sorts of information assessment task that is utilized to tie models depicting information classes or to expect future information plans [7]. Depiction measure has two stages; the first is the learning association where the coordinating instructive records are obliterated by amicable occasion assessment. The learned model or classifier is introduced as plan rules or models. The going with stage is the utilization of model for get-together, and test educational groupings are utilized to focus on the precision of depiction rules.

II. CLASSIFICATION

Approach is the way toward finding a model or an end that depicts and sees data classes and examinations, to use the model to expect the classes of things whose class mark isn't known. Data arrangements should be visible as a two-stage measure: learning step in which a classifier is made depicting a fated diagram of classes or encounters by separating the status set contained enlightening outline tuples and their associated names [4][5]. In the ensuing movement model is used for request by first surveying the reasonable precision of classifier worked during the basic turn of events. It is done using the test data. The accuracy of classifier on a given test set tuples is level of tuples that are effectively alluded to by the classifier. If the precision is over some sufficient level, the classifier can be used to expect future tuples whose class mark isn't known.

Portrayal is a kind of data evaluation that can be used to make models portraying huge data classes. System is a data mining approach used to predict pack pay for data models. It is one of the urgent plans in data mining and is used in various applications, for instance, plan affirmation, inconvenience confirmation, client relationship the pioneers, and directed

appearance. The goal of the portrayal assessments is to accumulate a model from a massive stack of getting ready data whose target class names are known and hence this model is used to pack covered cases [6] [8].

Plan is the most standard and most esteemed data mining systems. Structure maps data into predefined parties or classes. It is normal proposed as controlled getting the hang of thinking about how the classes are settled going before checking the data out. Methodology is the way toward finding a model that sees data classes, to use the model to expect the class of things whose class name is dull. The picked model relies upon the evaluation of a gigantic pile of orchestrating data. Edifying varieties are rich with camouflaged information that can be used for careful dynamic.

III. METHODOLOGY

This section gives the reduced idea about picked oversight models of Logistic Regression and Multilayered Perceptron.

3.1 Artificial Neural Networks (ANN)

ANN mirrors the cutoff points and exercises of the mind of person which is seen as the focuses, which is truly known as or called counterfeit neurons [4]. The neurons give and pass on information and data among themselves in sort of 0 s and 1 s or mix and every neuron has a particular weight given to it, which shows its capacities and undertakings to finish in the framework [5]. The advancement of ANN is separated into layers, clearly from information gathering layer, input layer, concentration or secret layer to yield layer which is called extraction or social occasion layer. Each layer has a particular capacity to perform and change information into the basic data to come by a definitive and ideal outcome. The Actuation and move work expect a fundamental part in the exercises do by neurons.

3.1.1 Multi-layered Perceptron (MLP)

A MLP is a holy person among the most in general saw Brain Organization plan that has been utilized for different applications. The MLP sort out is usually made from various fixations or administering units, and it is sorted out into a headway of no under two layers [7]. The focal layer (or the most decreased layer) is named as a data layer where it gets the outside data while the last layer (or the most astonishing layer) is a yield layer where the reaction for the issue is gotten. The mystery layer is the completely spellbinding layer in the information layer and the yield layer, and may graph with some spot very nearly one layers. The plan of MLP could be conceded as a nonlinear improvement issue. The target of MLP learning is to track down the best loads that limit the division between the data and the yield. The most unavoidable preparing evaluation utilized in NN is Back prompting (BP), and it has been utilized in directing different issues in model verification and depiction. This calculation depends a couple of limits, for example, striking covered focus fixations at the hidden layers learning rate, energy rate, endorsement work and the proportion of needing to occur.

3.2 Logistic Regression

Calculated Relapse on occasion called the determined model or logit model, analyzes the association between various free factors and a full scale subordinate variable, and assessments the probability of occasion of an event by fitting data to an essential twist [5]. There are two models of vital backslide, equal determined backslide and multinomial key backslide. Twofold essential backslide is consistently used when the dependent variable is dichotomous and the free factors are either steady or straight out. Exactly when the dependent variable isn't dichotomous and is involved past what two orders, a multinomial determined backslide can be used [6].

In a Strategic Relapse, the dependent variable simply has two characterizations. Generally, the occasion of the event is coded as 1 and its nonappearance as 0. Recollecting that codification changes the coefficients' sign and, along these lines, their impressive comprehension.[3][4] To all the more probable grasp how a determined backslide capabilities, understanding the reasoning of backslide examination overall is significant.

We should check the direct model's exemplary documentation: $Y = \alpha + \beta X + \epsilon$ out

Y addresses the reliant variable, that is to say, what we are attempting to comprehend/make sense of/anticipate. X addresses the free factor. The catch, (α), addresses the worth of Y when X equivalents zero. The relapse coefficient, (β), addresses the variety saw in Y related with the increment of one unit of X. The stochastic term, (ϵ), addresses the mistake of the model. In fact, it is feasible to gauge on the off chance that there is a straight connection between a reliant variable (Y) and different free factors. Besides, the model permits the perception of the impact size and to test the coefficients' measurable importance (p-worth and certainty spans).

IV. EXPERIMENTAL RESULTS

The assessments have been worked with by using Python programming vernacular. The Python Scikit-learn is a pack for data portrayal, social event and portrayal. We conducted experiments using two supervised learning algorithms: Artificial Neural Networks and Logistic Regression to build **machine learning models** that will detect if the mushroom is **edible or poisonous** by its specifications like cap shape, cap color, gill color, etc. using different classifiers. The dataset used in this paper is mushrooms. that contains 8124 instances of mushrooms with 23 features like cap-shape, cap-surface, cap-color, bruises, odor and two types of mushrooms are Eatable type contains 4208 and Poison type has 3916 instances [8]. Our experimental results demonstrate the following classification performance metrics for the selected machine learning algorithms as shown in the table-1 and figure-1.

Table-1
Experimental Results

| Algorithm | Accuracy | Precision | Recall |
|---------------------------|----------|-----------|--------|
| Artificial Neural Network | 96.47 | 96.5 | 96.5 |
| Logistic Regression | 94.56 | 94.5 | 94.5 |

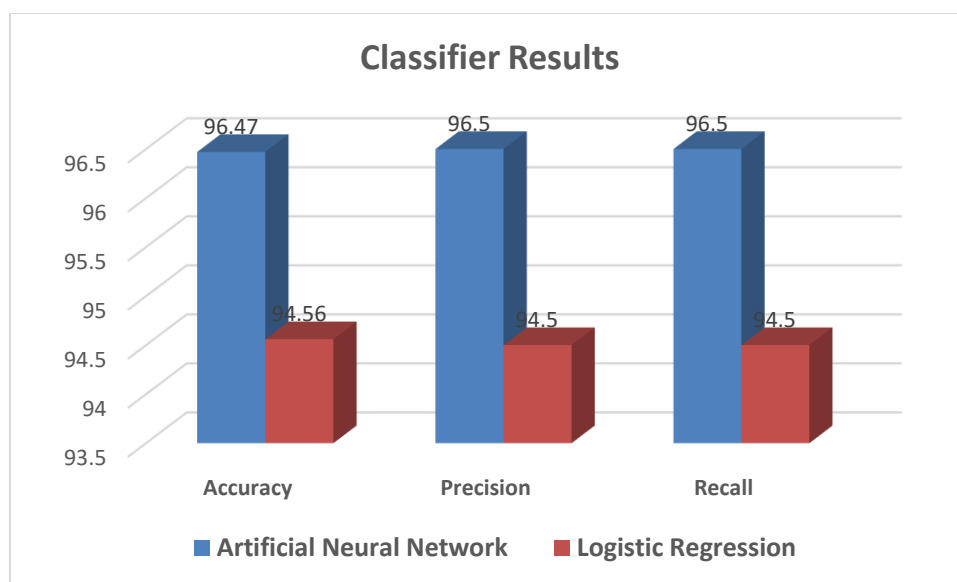


Figure-1: Performance of Classifier

From the figure-1, we observe the Artificial Neural Network outperformed Logistic Regression in terms of accuracy, precision, and recall, achieving an impressive accuracy rate of 96.47% in identifying the edibility of mushrooms. This suggests that the neural network model is highly effective in distinguishing between edible and poisonous mushrooms based on the provided features.

The research paper investigates the classification of mushrooms into edible or poisonous categories using machine learning algorithms. Two primary classifiers, Artificial Neural Network and Logistic Regression, were employed, and their performance was evaluated based on accuracy, precision, and recall.

Artificial Neural Network achieved an accuracy of 96.47%, with precision and recall both at 96.5%. This model outperformed Logistic Regression.

Logistic Regression yielded an accuracy of 94.56%, with precision and recall both at 94.5%.

4.1 Discussion

The results demonstrate the effectiveness of machine learning in accurately classifying mushrooms' edibility. The key points from the discussion are as follows:

Artificial Neural Network Superiority: The Artificial Neural Network proved to be the better classifier, achieving higher accuracy, precision, and recall compared to Logistic Regression. This suggests that complex models like neural networks can capture intricate patterns in mushroom features, leading to better discrimination between edible and poisonous mushrooms.

Practical Significance: Accurate mushroom classification is crucial for foragers, mycologists, and consumers to avoid ingesting toxic mushrooms. The high accuracy rate (96.47%) achieved by the Artificial Neural Network underscores its practical significance in promoting safe mushroom consumption.

V. CONCLUSION

In this research paper, we have presented an extensive analysis of machine learning models for the classification of mushrooms as either edible or poisonous. The results indicate that Artificial Neural Networks, with an accuracy of 96.47%, provide a robust and reliable method for mushroom classification. These models can play a crucial role in assisting foragers, mycologists, and consumers in making informed decisions regarding mushroom consumption.

The ability to accurately identify mushroom edibility is of paramount importance for both culinary enthusiasts and those who venture into the wild to harvest mushrooms. Misidentification of poisonous mushrooms can have severe health consequences, making the development of accurate classification models a valuable contribution to the field. Further research could explore the use of additional features or more advanced machine learning techniques to enhance the classification performance even further.

Overall, this study underscores the potential of machine learning in mushroom classification and its practical applications in ensuring the safety of mushroom consumption.

REFERENCES

- [1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)
- [2] G Ravi Kumar, K Tirupathaiiah and B Krishna Reddy, "[Client Churn prediction of banking and fund industry utilizing machine learning techniques](#)", IJCSE, Volume-7, Issue- 6, PP:842-846, 2019
- [3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [4] J.Han and M.Kamber, Data Mining concepts and Techniques, the Morgan Kaufmann series in Data Management Systems, 2nded.San Mateo, CA; Morgan Kaufmann, 2006.
- [5] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems", 2nd edition, Addison Wesley, 2005.
- [6] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.
- [7] Surya Bhupal Rao, S.Rahamat Basha, G. Ravi Kumar, "A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 120-131, ISSN: 0973-8975.
- [8] UCI machine learning repository. <http://archive.ics.uci.edu/ml/>.