➔ *Regular Research Paper – NS*

Classification Algorithms for Mushroom Identification

Sena GORAL
Mehmet Akif Ersoy University, Turkey
scelik@mehmetakif.edu.tr

Abstract

When examined around the world, it was determined that there are many types of mushrooms. Mushrooms are among the important nutrients for human life in many respects. An important feature of mushrooms used in various researches is that they are edible or poisonous. Different types of mushrooms have many characteristics. These features can be common in different features. However, the poisonousness of the mushroom can be determined by examining many features. In this study, whether the mushrooms are poisonous or edible is examined. In this study, 5 different classification methods were used. These; Logistic Regression, KNeigborsClassifier, DecisionTreeClassifier, RandomForestClassifier, and SVM. As a result, it was determined that SVM was the method that gave the most accurate result in this study.

Keywords: *Classification algorithm, mushroom, correlation matrix, machine learning, artificial intelligence.*

1. INTRODUCTION

Mushrooms; It is a food that can be picked by hand and has its own fruit parts (1). While some of the mushrooms are edible, some are used for medicinal purposes (2). Mushrooms are foods with high nutritional value. The amount of protein in it is very high. It can also be used as a supplement to increase immunity. In addition, mushrooms are used in the treatment of many diseases with their contents. It is important in every respect to use the right species of mushrooms, which have various uses and shapes.

There are many types of mushrooms in nature. Among these species, some are quite similar. Some species can be determined by testing in laboratories. For this reason, even experts can have difficulties in determining the mushroom species. It is important to ensure the sustainability of mushrooms, which are very important for humans, without being touched and harmed. In the data set used in the study, many features of fungi are stated. These features also vary within themselves. Depending on the properties, the classification of poisonous or edible changes.

Among the classification algorithms, the classification process of the fungi in the data set was carried out with 5 methods that are thought to be appropriate in the study. As a result of the successful results of the algorithms used, it is thought that the study will also be beneficial for researchers.

2. RELATED WORKS

Ortega et al. used 4 models for classification, namely logistic regression, naive bayes, decision tree and KNN. Decision tree classification technique got the most successful result with 88.2% [3]. Ottom et al. investigated whether mushrooms are poisonous or not with different techniques. Researchers have used Neural Network(NN), Support Vector Machines(SVM), Decision Tree and Nearest Neigbor(KNN) in their studies[4]. Masoudian and Mclsaac used the SVM method to detect fungal damage in their study. Over 90% accuracy has been achieved in their studies[5]. Subramaniam and Oh aimed to distinguish edible mushrooms from non-edible mushrooms in their study. The success rate of the method they recommend is between 85% and 96%. Achievement status changed in relation to training images[6]. Rahmat et al. proposed the k-nearest neigbor algorithm to identify poisonous mushrooms. After preprocessing the mushroom images, they started the identification process. The accuracy rate of the system was calculated as 90% for the 40 mushrooms they used to test the system[7]. Wibowo et al., using decision tree(C4.5), Naive Bayes and SVM algorithms, classified the mushroom data of Agaricus and Lepiota family. In the test results, 100% accuracy was determined for the C4.5 algorithm and SVM. However, it has been determined that the C4.5 algorithm gives faster results in terms of speed[8]. Maurya and Singh proposed the SVM classifier based on machine learning approach to distinguish between edible and poisonous mushrooms. The performance of the proposed approach was found to be 76.6%[9].

3. MATERIAL METHOD

In the data set used in the study, each mushroom has 23 different characteristics. One of these features is the mushroom's being poisonous or edible. The main purpose is to determine this feature of mushrooms. All values have been checked to ensure that the results of the data set are correct. Veil-type was found inappropriate in the review and was dropped. In addition, the cases of determining the poisonous or edible status of the properties are different. It has been deemed appropriate to use correlation matrix and heatmap technology to determine the features to be used for this purpose. The data set used is quite large. The variables in this data are cap-shape, cap-surface, cap-color, bruises, odor, gill-attachment, gill-spacing, gill-size, gill-color, stalk-shape ,stalk-root, stalk-surface-above-ring, stalk-surface-below-ring, stalk-color-above-ring, stalk-color-below-ring, veil-type, veil-color, ring-number, ring -type, spore-print-color, population, and habitat. Figure 1'de bazı özelliklerin kategori edilmesiyle oluşan sonuçlar paylaşılmıştır.

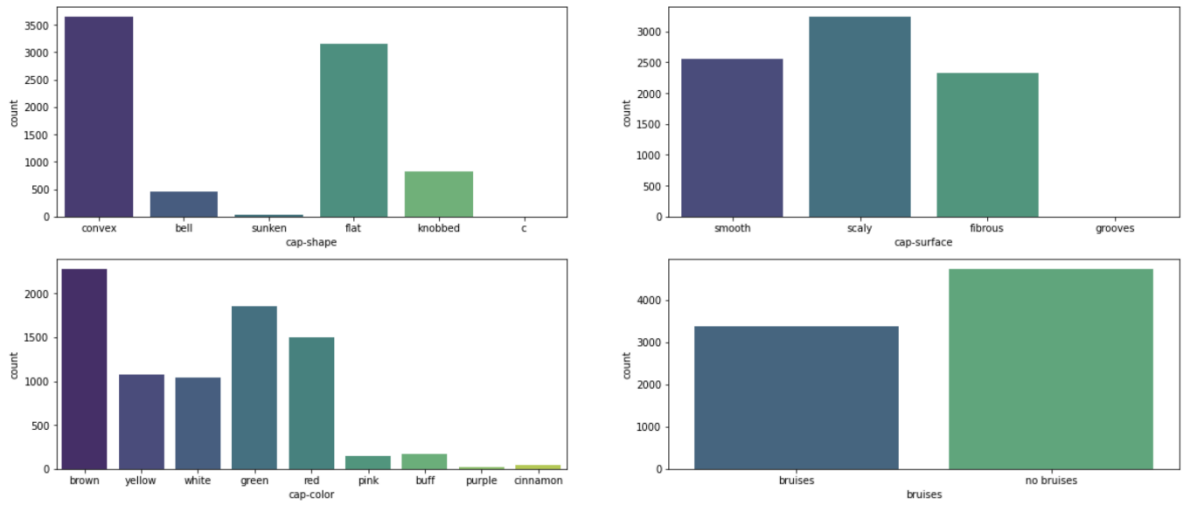

Figure 1. Distribution of some features

One of the 23 features specified in the data set is whether it is toxic or not. Figure 2 shows the distribution of mushrooms depending on whether they are poisonous or edible.

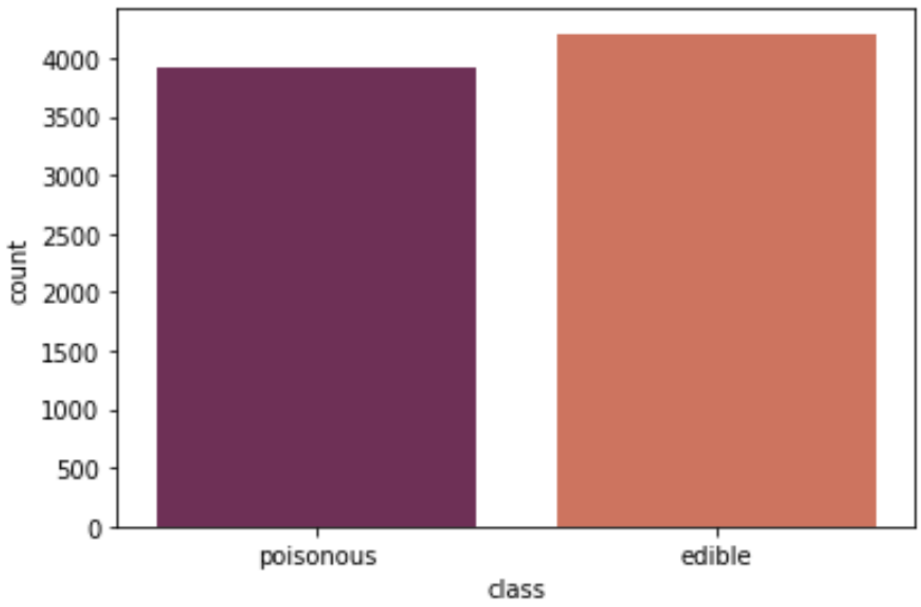

Figure 2. Poisonous and Edible Mushroom Count

All data have many different states within their properties. In this case, while some of the features are more effective for the decision, some of them are less effective. When the studies and researches are examined, the correlation matrix and heatmap are examined.

Necessary investigation was carried out for this study. As a result, the features to be included in the study were determined. These features are; cap-shape, cap-surface, bruises, gill-size, veil-color, spore-print-color, population. Correlation Heatmap of Mushroom Dataset Features was helpful in selecting these features. It is shown in Figure 3.

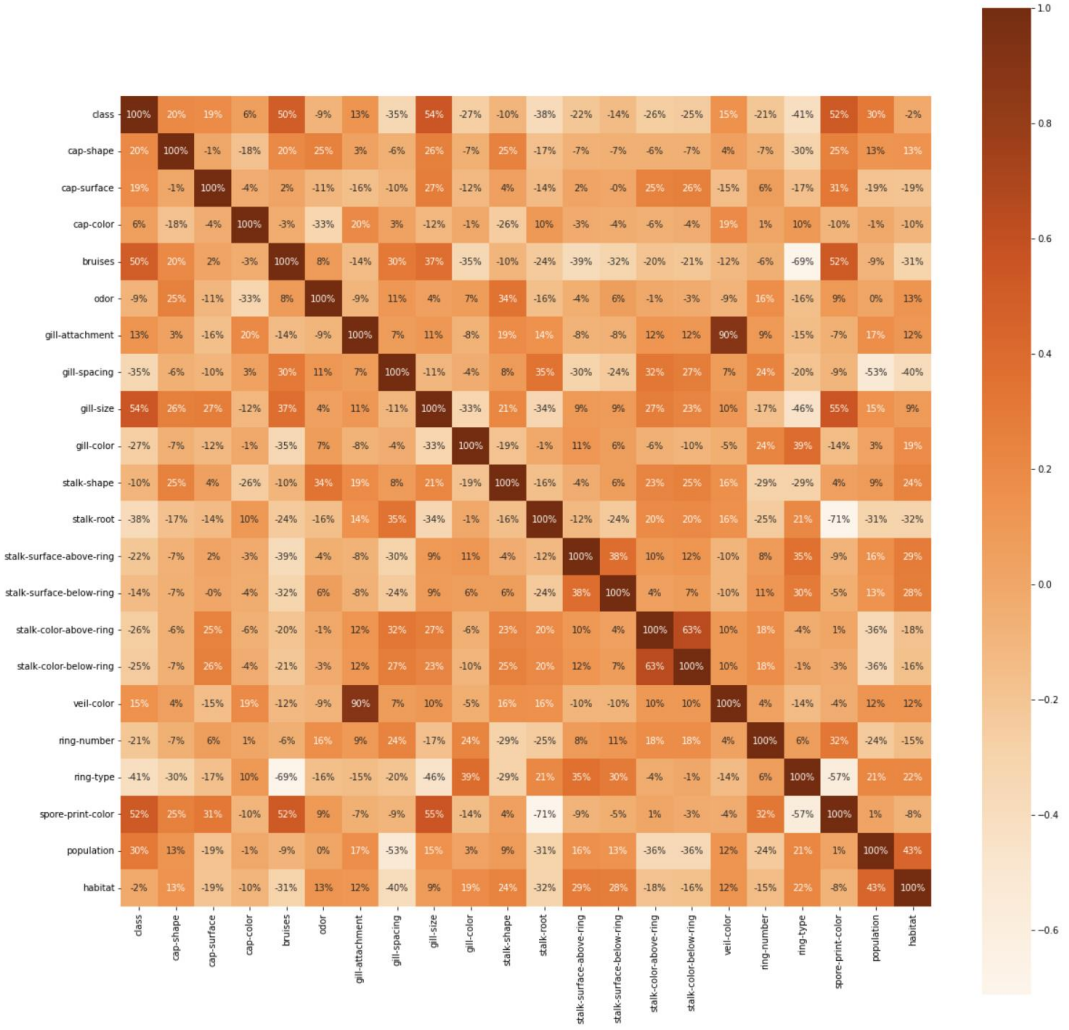

Figure 3. Correlation Heatmap of Mushroom Dataset

There are 8124 mushrooms and their properties in the data set. Classification was carried out using 5 different methods, namely Logistic Regression, KNeigborsClassifier, DecisionTreeClassifier, RandomForestClassifier and SVM. However, K-Fold Validation was applied for the data set in order to produce more accurate and robust results. Data separation is done with K-Fold Validation or Cross Validation. K-Fold Validation was used in this study. As a result, the split number was determined as 10 and operations were carried out. There are 8124 mushrooms and their properties in the data set. This dataset is divided into 813 mushrooms for test score and 7311 mushrooms for train score. 5

different models were evaluated for test score and train score. The results are shown in Table 1.

Table 1. Accuracy Score

Model	Train Score	Test Score
LogisticRegression	0.809574	0.803279
KNeighborsClassifier	0.987082	0.982216
DecisionTreeClassifier	0.990424	0.987688
RandomForestClassifier	0.990424	0.987688
SVM	0.976292	0.971272

The confusion matrix created for the data set reserved for the test is shown in Figure 4.

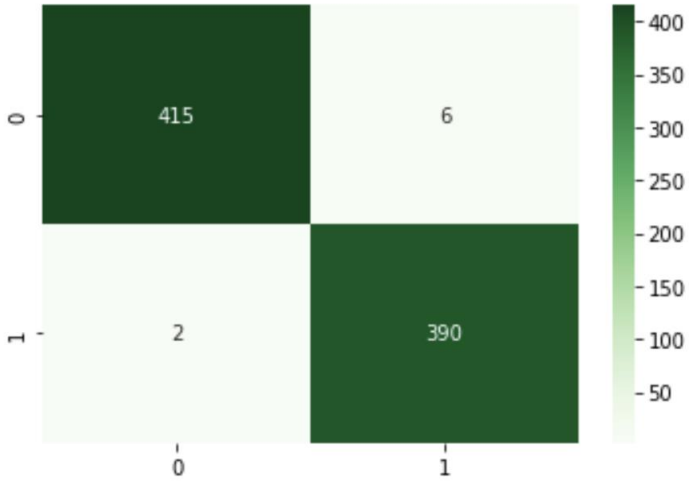

Figure 4. Confusion Matrix

4. CONCLUSION

The resulting classification report is shown in Table 2. The accuracy, precision, f1-score and recall values created here were found with the following calculations.

Accuracy=(TP+TN)/(TP+FP+FN+TN)

Precision=TP/(TP+FP)

Recall=TP/(TP+FN)

F1-Score=2X(PrecisionXRecall)/ (Precision+Recall)

Information:

TP = True positive;

TN = True Negative;

FP = False Positive;

FN = False Negative.

Table 2. Classification Matrix

	Precision	Recall	F1-Score	Support
0	1.00	0.99	0.99	421
1	0.98	0.99	0.99	392
Accuracy			0.99	813
Macro Avg	0.99	0.99	0.99	813
Micro Avg	0.99	0.99	0.99	813

It is thought that the study will be beneficial to researchers and interests. It is thought that the results will help people who do scientific studies with mushrooms. Compared 5 different classification methods are frequently used in other studies. Successful results were also obtained in this study.

REFERENCES

[1] Zhang, M., Cui, S. W., Cheung, P. C. K., Wang, Q. 2007. Antitumor polysaccharides from mushrooms: a review on their isolation process, structural characteristics and antitumor activity. Trends in Food Science & Technology, 18(1), 4- 19.

[2] Cheung, P. C. K. 2013. Mini-review on edible mushrooms as source of dietary fiber: Preparation and health benefits. Food Science and Human Wellness, 2, 162–166.

[3] Ortega, J. H. J. C., Lagman, A. C., Natividad, L. R. Q., Bantug, E. T., Resureccion, M. R., & Manalo, L. O. (2020). Analysis of Performance of Classification Algorithms in Mushroom Poisonous Detection using Confusion Matrix Analysis. International Journal, 9(1.3).

[4] Ottom, M. A., Alawad, N. A., & Nahar, K. M. (2019). Classification of mushroom fungi using machine learning techniques. International Journal of Advanced Trends in Computer Science and Engineering, 8(5), 2378-2385.

[5] Masoudian, A., Kenneth, A. M. 2013. Application of support vector machine to detect microbial spoilage of mushrooms. 2013 International Conference on Computer and Robot Vision, 28- 31 May, Regina, Canada, 281–287.

[6] Subramaniam, A., Oh, B. J. 2016 Mushroom Recognition Using PCA Algorithm. International Journal of Software Engineering and Its Applications, 10(1), 43-50.

[7] Rahmat, R. F., Aruan, T., Purnamawati, S., Faza, S., Lini T. Z. 2018. Fungus image identification using K-Nearest Neighbor. IOP Conference Series Materials Science and Engineering, 19-20 July, Medan, Indonesia, 420(1).

[8] Wibowo, A., Rahayu, Y., Riyanto, A., Hidayatulloh, T. 2018. Classification Algorithm for Edible Mushroom Identification. 2018 International Conference on Information and Communications Technology, 6-7 March, Yogyakarta, Indonesia.

[9] Maurya, P., Singh, N. P. 2020. Mushroom Classification Using Feature-Based Machine Learning Approach. Proceedings of 3rd International Conference on Computer Vision and Image Processing, 197-206.