Comparative Analysis of Interpretable Mushroom Classification using Several Machine Learning Models

Md. Sabbir Ahmed
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
msahmeds1997@gmail.com

Sadia Afrose
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
sadia.afrose@g.bracu.ac.bd

Ashik Adnan
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
ashik.adnan@g.bracu.ac.bd

Nazifa Khanom
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
nazifa.khanom@g.bracu.ac.bd

Md Sabbir Hossain
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd

Md Humaion Kabir Mehedi
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
humaion.kabir.mehedi@g.bracu.ac.bd

Annajiat Alim Rasel
Deaprtment of CSE
Brac University
Dhaka, Bangladesh
annajiat@gmail.com

Abstract—An excellent substitute for red meat, mushrooms are a rich, calorie-efficient source of protein, fiber, and antioxidants. Mushrooms may also be rich sources of potent medications. Therefore, it's important to classify edible and poisonous mushrooms. An interpretable system for the identification of mushrooms is being developed using machine learning methods and Explainable Artificial Intelligence (XAI) models. The Mushroom dataset from the UC Irvine Machine Learning Repository was the one utilized in this study. Among the six ML models, Decision Tree, Random Forest, and KNN performed flawlessly in this dataset, achieving 100% accuracy. Whereas, SVM had a 98% accuracy rate, compared to 95% for Logistic Regression and 93% for Naive Bayes. The two XAI models SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model Agnostic Explanation) were used to interpret the top three ML models.

Index Terms—Decision Tree, Explainable AI, Random Forest, KNN, SHAP, LIME

I. INTRODUCTION

Mushroom has the meaty substance of a fungus which bears lots of spores on its body. It is also known as toadstool and scientifically known as Agaricus bisporus. Mushroom belongs to the fungi family. Mushroom can be grown in anywhere and anytime. It can be grown from below ground to above the soil, and also from tropical to frosty places.

There are few types of mushrooms observed so far. Out of the estimated 1,500,000 species in the world, less than 69,000 species of mushrooms have been identified to date, and there are fewer than 200,000 species in Indonesia [1]. Those can be edible or medicinal, or even can be poisonous. When consumed, members of the deadly Agaricus and Lepiota families can make a person ill or perhaps kill them. You can

consume and even use wild members of the Agaricus and Lepiota families as medications [2].

Based on their growing place, those can be mild, tropical or subtropical. Mushroom recently refers to a robotic technique used in the food sector. This method was used to restrict attributes like color. The recent mushroom system uses particular qualities to enhance the selection of the process of mushrooms. A system like that depends on analysis and examining the features to improve the classification based on the recognized characteristics [3].

However, not many research works have discussed the interpretability of how a model is identifying an instance of mushroom as Edible or Poisonous. It's critical that humans can comprehend the conclusion or prediction produced by the AI or ML models, since it's a circumstance that might result in life or death. Explainable Artificial Intelligence (XAI) models can help to demonstrate the interpretability of any machine learning models as human understandable [4].

In order to classify mushrooms into those that are harmful to the body and those that are not, this study will examine the traits of each 23 species of gilled mushrooms in order to develop the various models and determine whether it is an edible mushroom or a poisonous mushroom and further show the intepretabbility of the models. The main focus of this study is to explain the reasons behind every models' behavior towards the classification of mushroom based on their features or traits by applying Explainable AI (XAI) models. SHAP (SHapley Additive exPlanations) [5] will be used for global explainability and LIME (Local Interpretable Model Agnostic Explanation) [6] for local or individual interpretability. This

work will also help to identify the key characteristics for the classification of Edible and Poisonous class.

The study is structured as follows: in part II, a quick overview of the related work, section III contains a description of the dataset and models, section IV provides the result and discussion of the proposed work which is followed by section V, which provides a conclusion.

II. Related Work

Based on a CNN model, mushroom is classified to be either edible or not in [1]. With an accuracy of 0.93, the proposed method of deep CNN or DCNN proves to be better in classifying mushroom. In static dataset, deep CNN or DCNN works better. It is also mentioned that since the performance of classification and network depth are unrelated, adding complexity won't improve the situation.

In another paper [7], it is found that the result of K-NN showed 100% accuracy rate. The reason for having the best result is simply for the dataset which was numeric data with discrete value. This kind of dataset is highly suitable for K-NN algorithm, which is why the accuracy rate is this high.

Again, based on the results of evaluating the top three data mining classification techniques, comparative classification algorithm testing accuracy in prior data mining has not been conducted. In comparison to the other two widely used classification algorithms, the C4.5 algorithm is the most accurate and has the fastest processing times. This approach produces a decision tree that may be readily used to build applications [2].

In [3], they used a variety of machine-learning classifiers on a dataset called "mushroom data" that was provided, and the dataset is from UCI repository. They discovered that one characteristic, "stalk-root," has numerous missing values, while another, "veil-type," has identical values across all rows. In order to avoid their impact on the classifications, these two traits are removed. Whereas, the "odor n" feature is the element that influences decisions the most. Therefore, according to the paper, a mushroom is more likely to be inedible if it has an odor. Due to the data set's cleanliness, the majority of classifiers appear to perform well on it. Nevertheless, the decision tree, ANN, and SVM classifiers perform better than the other classifiers. Moreover, in order to increase accuracy, they proposed a hybrid model that combines the most effective classifiers.

Here in this paper [8], the attempts were taken to improve the results by removing the background from the photographs, but the effort was unsuccessful. The experiment's findings also indicate that background images have an advantage, particularly when using the KNN algorithm, Eigen features extraction, and the real dimensions of mushrooms, where accuracy reached 0.944, while the result when the real dimensions are replaced with virtual dimensions is 87%. After removing the background of the photos, the KNN value achieved a maximum of 0.819. In our upcoming work and also according to the paper, certain physical characteristics of mushrooms, such as cup sizes, stem heights, color, and textures can be used for getting better outcome.

The accuracy of the four ML models used in the research [9] is 90.99%, 98.82%, 99.98%, and 100% for Naive Bayes, Decision Trees, Support Vector Machines, and AdaBoost algorithms, respectively. However, there was a lack of explanation in the work. Another study [10] applied Naive Bayes, Decision Tree (C4.5), Support Vector Machine (SVM), and Logistic Regression and came to the conclusion that the c4.5 method shows a maximum accuracy of 93.34%. K-Nearest Neighbor and Decision Tree models were compared by Chitayae et al. (2020), who came to the conclusion that Decision Tree performed better [11]. In several articles, deep learning-based methodologies were applied, however they also failed to demonstrate the explainability of their models [12] [13].

III. Methodology

A. Dataset Description and Preprocessing

The Audubon Society Field Guide to North American Mushrooms' descriptions of hypothetical samples for 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom are included in this dataset [14]. Every species is labeled as either unquestionably edible, definitely poisonous, or of unknown edibility and not advised. The last two classes were merged as a single class.

The dataset has 8124 rows and 23 columns, and there are 4208 instances of the Edible class and 3916 instances of the Poisonous class. The dataset is therefore balanced. Since the information is categorical, we will use LabelEncoder to make it ordinal. The LabelEncoder transforms each value in a column into a number. The category column must be of the data type "category" in order to use this method. A non-numerical column's datatype by default is "object." Following the conversion of the columns to the "category" type, LabelEncoder was used to make the ordinal conversion. The column "veil type" was also removed because it has a value of 0 and does not add anything to the data.

B. Model Description

Decision Tree: It's a different-supervised regression and classification machine learning approach. Decision trees can be used at classification task, training models to predict the target variable's class or value from prior data (training data). Decision Trees compare root and record attribute values and predict the next node's branch from the tree's root.

Support Vector Machine (SVM): SVM is a supervised technique for classification and regression. SVM classifies

future data points by selecting the optimal line or decision boundary, which helps partition n-dimensional space into classes.

Random Forest: The random forest consists of multiple decision trees. It employs bagging and feature randomization to produce an uncorrelated forest of trees whose prediction by committee is more accurate than any individual tree. This approach classifies continuous and categorical variables better than it does regression.

Logistic Regression: Logistic regression predicts the likelihood of a target variable using supervised learning classifications. A dichotomous target or dependent variable has only two classifications. Logistic regression is categorical and discrete or categorical results are required here.

K-Nearest Neighbor (KNN): It is a supervised learning classifier which is non-parametric and employs closeness to group data points. It can be used for regression or classification, although it's usually a classification approach that assumes comparable points are close together.

Naive Bayes: It is a Bayes Theorem-based machine learning algorithm used for categorization. A Naive Bayes classifier presupposes that a class's features are unrelated. It's been used for many purposes, but it excels at NLP.

IV. RESULTS AND DISCUSSION

Three of the six ML models that we employed for the classification task performed exceptionally well, and among those three, two of them had 100% accuracy rate. The other three models did quite good as well.

From Table-1, We can observe that, in case of Support Vector Machine, the Accuracy, Precision, Recall and F1 Score are respectively 98%, 97%, 98% and 98% for the Edible class and 98%, 98%, 97% and 97% respectively for the Poisonous class. In case of Logistic Regression, the Accuracy for both Edible and Poisonous class is 95%, Precision for Edible class is 96% and Poisonous class is 94%, Recall is respectively 94% and 96% for Edible and Poisonous class and the F1 Score is 95% for both the classes. The worst performing model among the proposed models is Naive Bayes with 93% Accuracy for both Edible and Poisonous class. The Precision, Recall and F1 Score for the Edible class are respectively 94%, 92% and 93%. For the Poisonous class, the Precision, Recall and F1 Score are 91%, 94% and 92% respectively. On the other hand, our second best performing model is the K-Nearest Neighbor (KNN) with 100% Accuracy and F1 Score for both classes and 100% Precision and 99% Recall for the Edible class and 99% Precision and 100% Recall for the Poisonous class. Finally, the best two models are Decision Tree and Random Forest. For both models, we can observe that the Accuracy, Precision,

Recall and F1 Score of both Edible and Poisonous class are 100%.

We will go into depth on the interpretability of the top three models for identifying the edible and poisonous classes using two XAI models, SHAP and LIME. By calculating the Shapely values for the entire dataset and combining them, SHAP can determine the global interpretation.
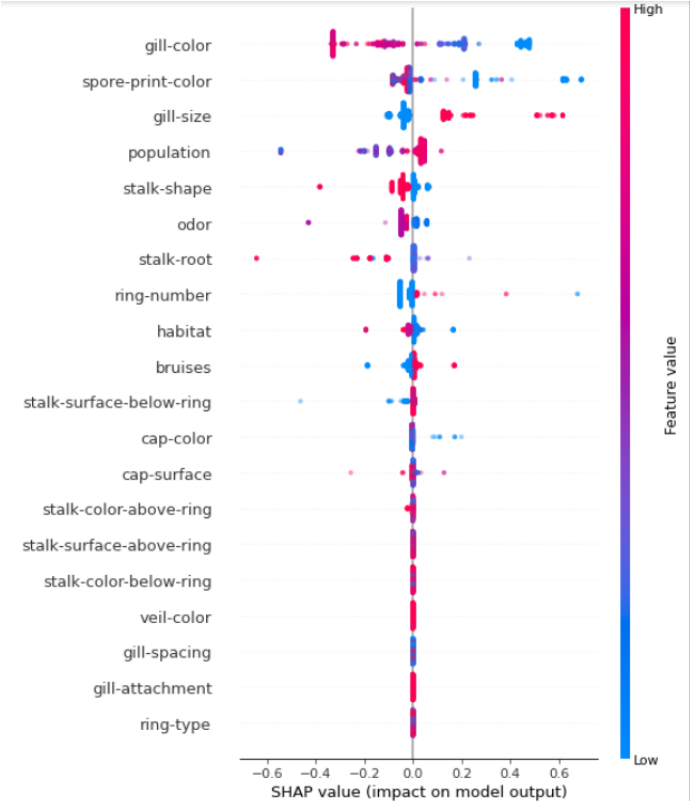

Fig. 1. Summary of SHAP values over all features for Decision Tree

Fig. 1 shows a global summary of the SHAP values distributed over the features of the Decision Tree model. For each feature, the distribution of feature importance can be observed. From Fig. 1 it can be seen that gill color, spore print color, gill size, population, stalk shape, and odor have the greatest effects on the prediction over the whole dataset. The high SHAP value are shown on the x-axis. High gill color, spore print color, stalk shape, and odor values impact the prediction negatively as red values are on the left-hand side for these features, while high gill size and population values affect the prediction in a positive way as red values are on the right-hand side. In a similar manner, low gill color, spore print color, stalk shape, and odor values affect the prediction positively, and low gill size and population values affect the prediction negatively.

An overview of the SHAP value distribution across all characteristics for the Random Forest model is shown in Fig. 2. The distribution of feature importance is shown for each feature (horizontal rows). As seen in the diagram, odor, gill

|	Edible				Poisonous			
Model Name	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Decision Tree	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Support Vector Machine (SVM)	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.97
Random Forest	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Logistic Regression	0.95	0.96	0.94	0.95	0.95	0.94	0.96	0.95
K-Nearest Neighbour (KNN)	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**
Naive Bayes	0.93	0.94	0.92	0.93	0.93	0.91	0.94	0.92
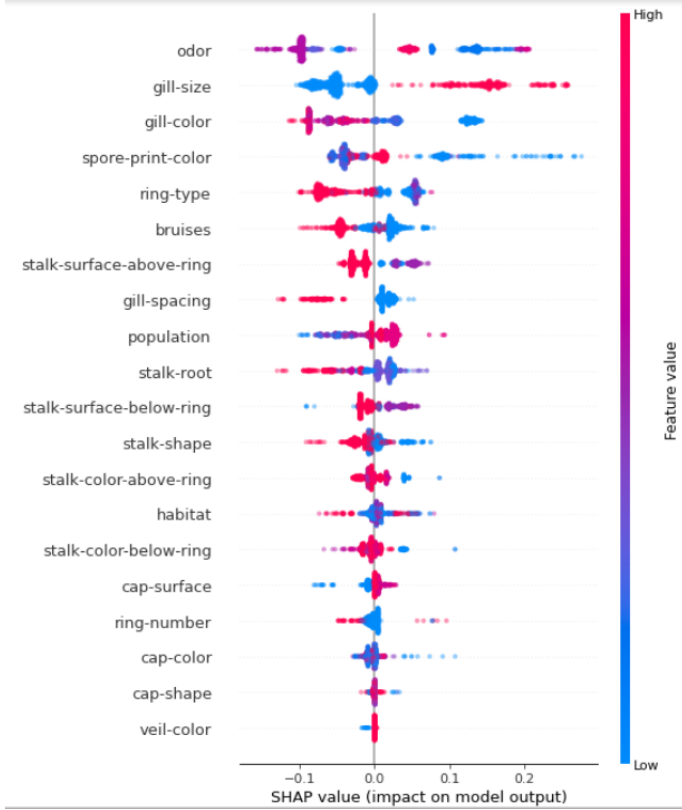
Fig. 2. Summary of SHAP values over all features for Random Forest
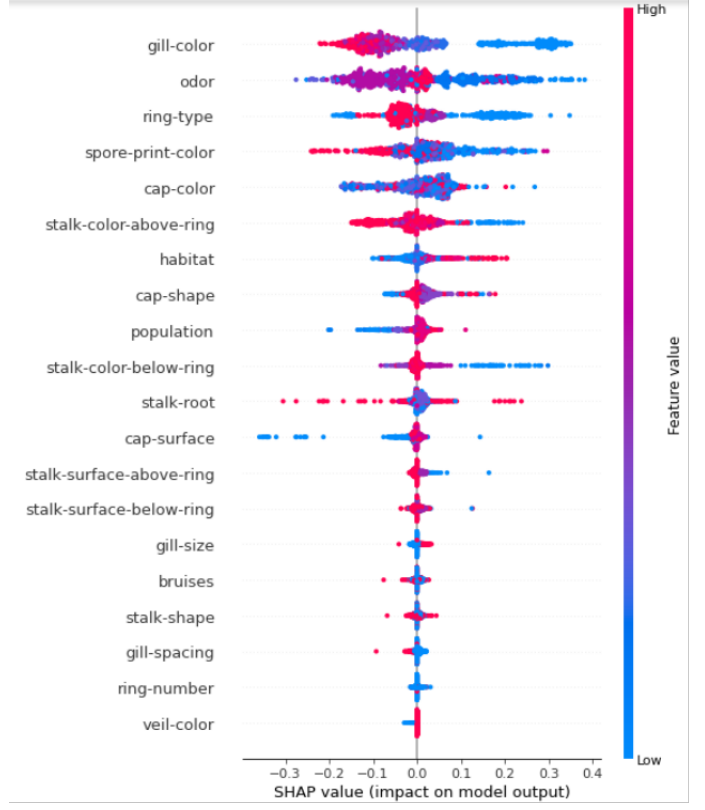

Fig. 3. Summary of SHAP values over all features for K-Nearest Neighbour (KNN)

color, gill size, spore print color, ring type, and bruises have a significant impact on prediction over the whole dataset. High gill color, ring type, and bruises values have a negative impact on the prediction as red values are on the left, but high odor, gill size, and spore print color values have a positive effect as red values are on the right-hand side. In a similar manner, low gill color, ring type, and bruises values have a positive impact on the prediction, whereas low odor, gill size, and spore print color values have a negative influence.

A global summary of the SHAP value distribution across all characteristics for the K-Nearest Neighbour (KNN) model is shown in Fig. 3. The distribution of feature importance is displayed for each feature (horizontal rows). We can observe from the diagram that gill color, odor, ring type, spore print color, cap color, stalk color above ring and

habitat all have a significant impact on the prediction over the whole dataset. High gill color, odor, ring type, spore print color, and stalk color above ring (mostly) values affect the prediction in negative way as red values are on the left-hand side, high habitat values, on the other hand, have a favorable impact on the forecast since the red values are on the right. Similarly, low gill color, odor, ring type, spore print color, and stalk color above ring (mostly) values affect the prediction positively and low habitat values affect the prediction negatively. Cap color, however, does not exhibit a pronounced separation of importance from the other top features.

The three best models for our proposed classification system have been discussed in terms of their SHAP values,

and it is clear that key features, such as gill color, spore print color, and odor, are shared by all three models as being among their most significant features. Therefore, we may infer that these features are the most crucial characteristics for classifying edible and poisonous mushrooms.

LIME offers local model interpretability in contrast to SHAP. LIME adjusts one data sample by changing the feature values, then tracks the effect on the output. This frequently relates to the questions that people ask while looking at a model's results. The next section will discuss LIME on our top three ML models.
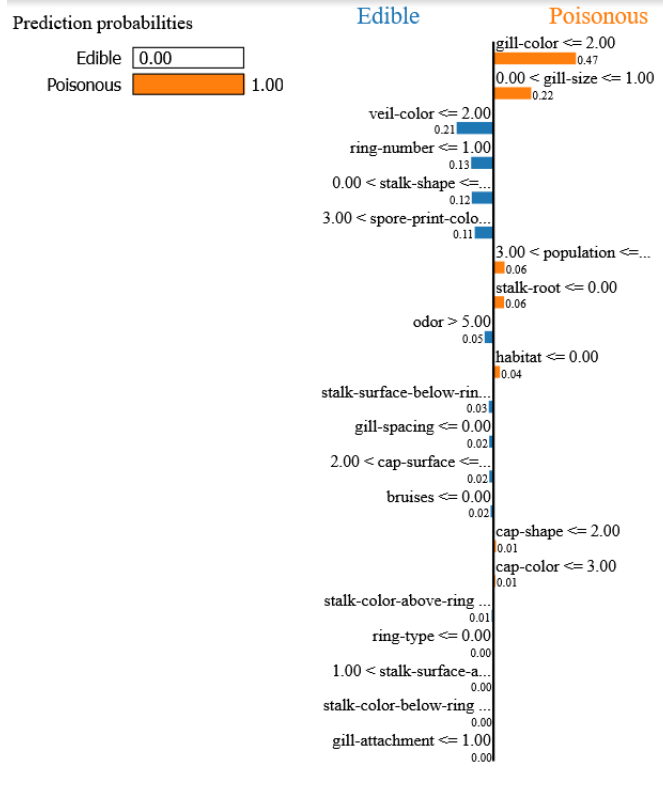

Fig. 4. LIME explainability for a single instance of the Decision Tree model

From Fig. 4, it can be observed that in case of Decision Tree, the model predicts this particular instance as poisonous with 100% confidence and explains the prediction because of the features gill color, gill size, population, stalk root, habitat, cap shape, and cap color. All of these features converge toward class 1, which is the poisonous class. Out of all of these features, gill size and gill color had the most influence on the prediction towards the poisonous class.

From Fig. 5, it is clear that for Random Forest, the model accurately predicts that this specific instance is poisonous with 100% confidence and provides an explanation for the prediction based on the features of the sample, including odor, gill size, gill spacing, gill color, ring type, population, stalk color below ring, bruises, stalk color above ring, habitat,
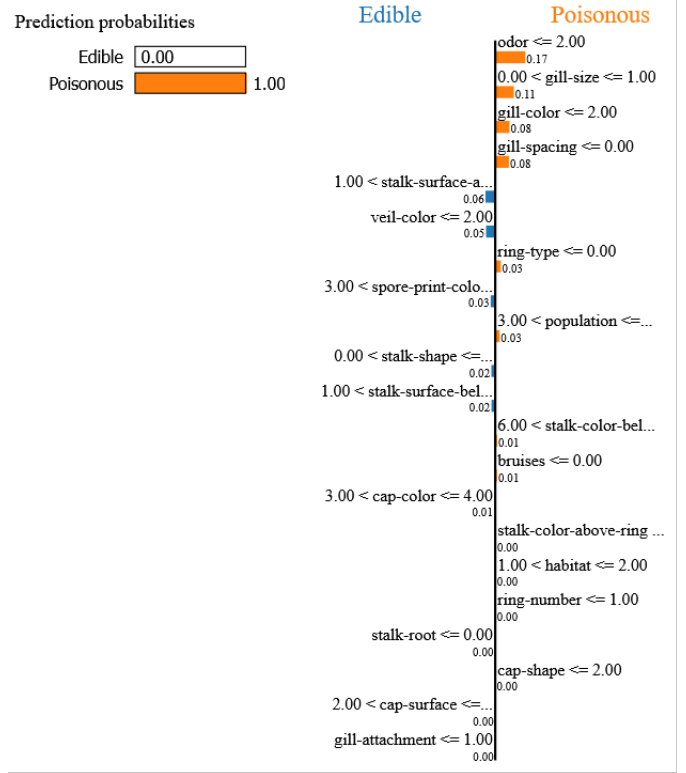

Fig. 5. LIME explainability for a single instance of the Random Forest model

ring number, and cap shape. These features all go toward class 1, which is the poisonous class. Odor, gill size, color, and spacing had the most impact on the prediction of the poisonous class out of all of these features.

Inferred from Fig. 6 is that the model predicts with 100% confidence that this particular instance is poisonous, and offers an explanation for the prediction based on practically all of the features for K-Nearest Neighbour (KNN). These features are all indicative of class 1, which is the poisonous class. But out of all of these features, odor, veil color, spore print color, and habitat had the greatest impact on the classification of the poisonous class.

V. CONCLUSION

The objective of our proposed research was to demonstrate the interpretability of machine learning models for classifying edible and poisonous mushrooms. In order to do that, the two XAI models, SHAP and LIME, were employed. We demonstrated the global and local interpretability of the top three ML models from our proposed work, and as a result, we obtained some crucial knowledge about the best features for the classification task. The combination of various XAI models with Deep Learning and Machine Learning models may be a potential future research work.
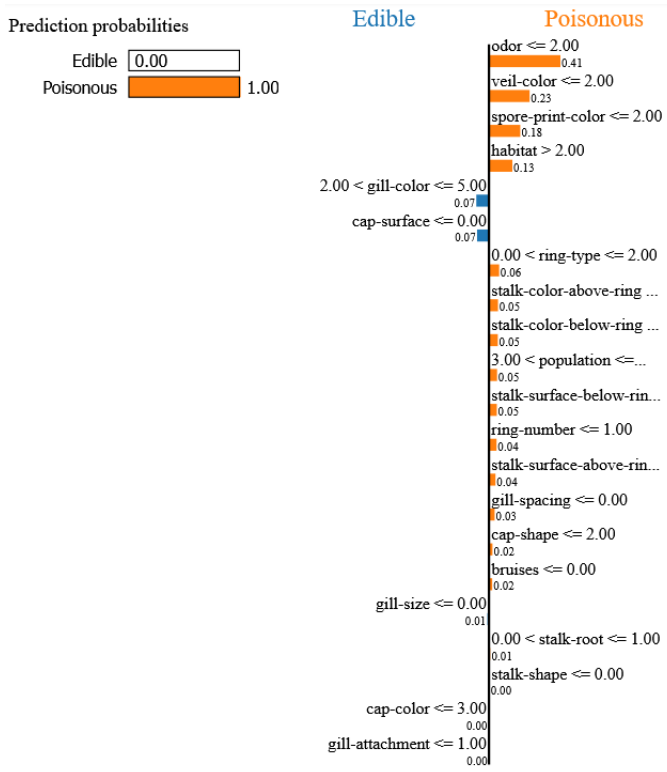
Fig. 6. LIME explainability for a single instance of the K-Nearest Neighbour (KNN) model

REFERENCES

[1] G. Devika and A. G. Karegowda, "Identification of edible and non-edible mushroom through convolution neural network," in *3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021)*. Atlantis Press, 2021, pp. 312–321.

[2] A. Wibowo, Y. Rahayu, A. Riyanto, and T. Hidayatulloh, "Classification algorithm for edible mushroom identification," in *2018 International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2018, pp. 250–253.

[3] O. Tarawneh, M. Tarawneh, Y. Sharrab, and M. Altarawneh, "Mushroom classification using machine-learning techniques," 07 2022.

[4] D. Gunning and D. Aha, "Darpa's explainable artificial intelligence (xai) program," *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.

[5] M. Chromik, "reshape: A framework for interactive explanations in xai based on shap," in *Proceedings of 18th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET), 2020.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016.

[7] N. Chumuang, K. Sukkanchana, M. Ketcham, W. Yimyam, J. Chalermdit, N. Wittayakhom, and P. Pramkeaw, "Mushroom classification by physical characteristics by technique of k-nearest neighbor," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*. IEEE, 2020, pp. 1–6.

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

[9] K. Tutuncu, I. Cinar, R. Kursun, and M. Koklu, "Edible and poisonous mushrooms classification by machine learning algorithms," in *2022 11th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 2022, pp. 1–4.

[10] K. Kousalya, B. Krishnakumar, S. Boomika, N. Dharati, and N. Hemavathy, "Edible mushroom identification using machine learning," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2022, pp. 1–7.

[11] N. Chitayae and A. Sunyoto, "Performance comparison of mushroom types classification using k-nearest neighbor method and decision tree method," in *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2020, pp. 308–313.

[12] N. Zahan, M. Z. Hasan, M. A. Malek, and S. S. Reya, "A deep learning-based approach for edible, inedible and poisonous mushroom classification," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2021, pp. 440–444.

[13] N. Kiss and L. Czúni, "Mushroom image classification with cnns: A case-study of different learning strategies," in *2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2021, pp. 165–170.

[14] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml