COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS TO IDENTIFY EDIBLE AND POISONOUS MUSHROOMS

ABSTRACT

Mushrooms are actually fungi kingdom, but comes under the category of vegetables for cooking purposes. They are fat-free, low-sodium, low-calorie, and cholesterol-free. Mushrooms are of different types both eatable and poisonous, each mushrooms have unique look and taste. But nutritional benefits vary depending on the type of mushroom. They contain essential nutrients such as proteins, vitamins, minerals, amino acids, antibiotics and antioxidants. Mushrooms are good for our health. But all species of mushrooms are not edible, some of them are poisonous and it may cause health problems and leads us to death. So before eating, it should be checked whether it is edible or not. Actual determination and proper identification of species are the only safe way to ensure edibility and safeguard against possible accident of consuming poisonous one . This paper is a comparison of different machine learning techniques on mushroom dataset to identify which algorithm is best for the identification of edible and poisonous mushrooms.

INTRODUCTION

There are thousands of species of mushrooms in the world; they are edible and non-edible being poisonous. It is difficult for non-expertise person to Identify poisonous and edible mushroom of all the species manually. So a computer aided system with software or algorithm is required to classify poisonous and nonpoisonous mushrooms. Consumption of poisonous mushrooms may vary from gastrointestinal upset to life-threatening organ failure resulting in death. Therefore care must be taken to classify edible mushrooms with that of poisonous ones. Mushroom classification is done with the help of ***weka tool*** (Waikato Environment for Knowledge Analysis).

TOOL USED

WEKA (Waikato Environment for Knowledge Analysis), a tool for data pre-processing, classification, regression, clustering, association rules, and visualization is made use of for the prediction. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. This paper aims in identifying edible and poisonous mushrooms based on certain attributes like shape, size, color, etc. Here we consider the mushroom dataset which contains many attributes. Apart from these attributes, we can also include other attributes like cap margins, cap size, stem color, ecology, protein content, toxins, taste etc. Here we apply different machine learning techniques on mushroom dataset and is evaluated using its accuracy, mean absolute error, kappa statistics etc. Based on that evaluation, we can easily identify which algorithm is best for the identification of edible and poisonous mushrooms.

INPUT VARIABLES

#	Input Variable	Domain
1	cap-shape	bell=b , conical=c , convex=x , flat=f , knobbed=k , sunken=s
2	cap-surface:	fibrous=f , grooves=g ,scaly=y , smooth=s
3	cap-color:	brown=n, buff=b, cinnamon=c , gray=g , green=r , pink=p , purple=u , red=e white=w , yellow=y
4	bruises?:	bruises=t ,no=f
5	odor:	almond=a ,anise=l , creosote=c , fishy=y , foul=f , musty=m , none=n , pungent=p , spicy=s
6	gill-attachment:	attached=a ,descending=d ,free=f ,notched=n
7	gill-spacing	close=c ,crowded=w ,distant=d
8	gill-size:	broad=b ,narrow=n
9	gill-color:	black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
10	stalk-shape:	enlarging=e , tapering=t
11	stalk-root:	bulbous=b ,club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?
12	stalk-surface-above-ring:	ibrous=f, scaly=y ,silky=k, smooth=s
13	stalk-surface-below-ring:	ibrous=f, scaly=y, silky=k, smooth=s
14	stalk-color-above-ring:	brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
15	stalk-color-below-ring:	brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
16	veil-type:	partial=p, universal=u
17	veil-color:	brown=n, orange=o, white=w, yellow=y
18	ring-number:	none=n, one=o, two=t
19	ring-type:	cobwebby=c, evanescent=e, flaring=f, large=l, none=n ,pendant=p, sheathing=s, zone=z
20	spore-print-color:	black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y