Magic mushrooms also called shrooms are hallucinogenic psychoactive substances that can distort perception of time and objects. Psilocybin and psilocin are the main compounds in shrooms that cause hallucinations. If taken orally which Is the usual case, psilocybin will be converted into psilocin which will be the substance that initiates the hallucinations. Eating shrooms are the typical method but can also be brewed in tea which also reduces bitterness. One major concern is not all shrooms are edible and there are some poisonous shrooms that can be fatal. People who decide to consume shrooms should have some prior knowledge of what to look for to reduce the risk of eating a poisonous shroom and possibly death. The main aim of this research is to explore if edible and poisonous shrooms can be distinguishable by certain features. It is hypothesised edible and poisonous shrooms can be distinguished with high accuracy by machine learning.

**Method**

The dataset topic was shroom classification and was downloaded from Kaggle. Python 3 was used to explore and analyse the data. The main packages used to run the analysis was Pandas, Sklearn and matplotlib. There were 8124 samples and there were no missing values.

**Features**

The dataset consisted of 22 features describing different aspects of shrooms and all features were categorical. A few of the features were bruises, odour, cap-shape, gill attachment, population, and habitat. Pandas get dummies method was used on all features so feature labels were converted to either a 0 or 1. This increased the feature count from 22 to 95.

**Target**

The target was shroom class and had two labels: edible and poisonous. Sklearn Label Encoder method was used to convert the labels to integers. For example, edible was converted to 0 and poisonous was converted to 1. The target was also balanced with 4208 samples consisting of edible and 3916 samples consisting of poisonous.

**Machine learning models**

Various machine learning models were used to analyse the data including logistic regression, decision tree, random forest, and a multi-layer perceptron.

**Logistic regression**

6 out of the original 22 features were used in the first logistic regression model. After using the get dummies method, there were now 28 features. Sklearn Logistic Regression class and train\_test\_split function was used to run the analysis. There were 5686 samples in the training set and 2438 samples in the test set. The logistic regression fit function was used to fit the training data and the predict function was used to evaluate model performance on the test set. A second logistic regression model was used, and the exact same steps were used except all features were used this time.

**Decision tree**

Sklearn DecisionTreeClassifier class and GridSearchCV function were used to run the analysis. A dictionary was created to test various values of 3 parameters in the decision tree model. Various values of maximum tree depth, minimum samples for each leaf node and maximum number of leaf nodes were tested using grid search function and 10-fold cross validation was used. The training and testing data was the same proportion as the logistic regression models. The decision tree classifier was then created with a maximum depth of 15, maximum leaf nodes as 20 and minimum samples per leaf node as 1 as these parameter values were considered the best from the grid search. The Decision tree classifier fit function was then used on the training data and the predict function was used to evaluate model performance.

**Random Forest**

Sklearn RandomForestClassifier class and GridSearchCV function was used to the run the analysis. A dictionary was created to test various number of decision trees to build in the random forest model ranging from 1-100, and 10-fold cross validation was used. The training and test data was the same proportion as the previous models. A plot was graphed with decision tree number on the x axis and accuracy on the y axis. The number of trees to build the random forest model was set as 5 considering after this value, the accuracy levelled off. The random forest model was then fitted on the training data and predict function was used to evaluate model performance.

**Neural network**

Keras API was used to build the multi-layer perceptron and the model was created using the sequential model. The first layer was a fully connected layer with 10 nodes, input size of 95 and relu activation function was used. The second layer was also a fully connected layer and consisted of 20 nodes and the relu activation function also used. The output layer consisted of 1 node and the sigmoid activation function was used. The model was then compiled with Adam optimizer (learning rate 0.0001) and the loss function was binary cross entropy. The model was then fitted with the training data, 20% of the training data was used for validation split, batch size was 32 and there were 100 epochs. The Sequential API evaluate function was used to then evaluate model performance on the test set.

**Results**

*Table 1*. Test set accuracy for each machine learning model.

|  |  |
| --- | --- |
| **Model** | **Accuracy** |
| Logistic regression (6 features) | 99 % |
| Logistic regression (all features) | 100 % |
| Decision tree | 100 % |
| Random forest | 100 % |
| Neural network | 100 % |

As can be seen from table 1, the logistic regression model using a subset of the features resulted in test set accuracy of 99 %. The logistic regression model with all features, decision tree, random forest and neural network models all scored 100% classification accuracy on the test set.

**Discussion**

The hypothesis that edible and poisonous shrooms can be classified with high accuracy was supported. 4 out of the 5 machine learning models scored 100% classification accuracy on the test set while the logistic regression model with a subset of features was still able to score 99 %. Therefore, machine learning can be an invaluable tool to distinguish edible and poisonous shrooms with very high accuracy. More specifically, even simple machine learning models such as logistic regression performed just as good as more complex models such as random forest and multi-layer perceptron. The major implication of the research is machine learning can greatly reduce the risk of people consuming poisonous shrooms that may possibly cause death. In conclusion, machine learning can be an invaluable tool to greatly reduce the risk of people consuming poisonous shrooms and therefore making the experience safer.