Mushroom Classification

Abstract:

Mushrooms are an important fungus which contains a good source of vitamin B and a large amount of protein when compared to all other vegetables. It helps to prevent cancer; it is useful in weight loss and increases the immunity power in humans. On the other hand, some mushrooms are toxic and can be dangerous. Therefore, it is a prominent task to differentiate the edible and poisonous mushrooms.

Project Scope:

- Parse the data (preprocessing the data)
- Build a machine learning model that will predict whether a mushroom is edible or not, based on the other characteristics.

Design Steps/Methodology:

1. Data Preparation:
- Load the dataset.
- Understand the data.
- Preprocess the data by handling missing values, encoding categorial variables and performing the required transformations.
2. Feature Selection and Engineering:
- Identify those features that can predict the eligibility of mushrooms.
- Engineer new features, if necessary, based on domain knowledge or data insights.
3. Model Selection and Training:
- Split the dataset into training and testing sets to evaluate the performance of model.
- Choosing the machine learning algorithm for binary classification. Ex: - Logistic Regression, Decision trees, Random Forest classifiers.
- Train the selected model on training data and tune hyperparameters using techniques like cross –validation or grid search.
4. Model Evaluation:
- Evaluate performance of trained model using evaluation metrics like accuracy, precision, recall and F1 score.
- Analyze any class imbalance issues and consider using techniques like oversampling and under sampling to address the imbalance

Result and Outcomes:

The methods that we used here are: -

- Support vector machine = 100% accuracy

- Decision Tree Classifier = 97.9% accuracy
- Logistic Regression = 95% accuracy

Benefits:

Classifying mushrooms can have several benefits, especially when it comes to distinguishing edible mushrooms from poisonous ones.

The key benefits include: -

1. Food Safety
2. Culinary Applications
3. Medicinal Uses
4. Biodiversity Conservation
5. Scientific Research
6. Recreational and Educational Purposes

Accurate identification plays a crucial role in maximizing the benefits and minimizing the risks associated with mushrooms.

Lessons Learnt:

1. Understanding the problem: - It took some time for me to understand the main purpose of this particular project. After understanding the objectives, requirements and constraints, I could easily be able to classify the information.
2. Data preprocessing: - I learnt that this is a very critical stage in handling any data. Understanding the variables and statistical quality of the given data, parsing the data from csv file into readable format. I have put most of my time to clean the data, clearing the null values, formatting the values, replacing the missing values and encoding the data to binary values.
3. Model selection: - There are different types of methods for encoding the data. In the selection of machine learning algorithm, I have used three methods and those being Logistic regression, Decision Tree Classifier, and Support Vector Machine (SVM).
4. Model Interpretation and Error analysis: - Splitting the data into training and testing was a time-taking task for me, as this part was completely new, and I studied how it needs to be done and the underlying process for it. I did face some challenges while coding and ran into errors which I couldn't solve. Then resolved those with the help of my colleagues and Stack overflow.
5. Continuous learning: - As I have an interest in AI & ML, I was overwhelmed to get this project. I was keen to know every detail of this project and I went through most of the documentation and various forums as this was my first project.

Conclusion:

The accuracy for the methods used is more than 95% which means what we have is edible. Most of the classifiers perform well as the dataset is a clean set.