**Sub-questions for Problem 2:**

1.)Why did you choose the particular algorithm?

In the example provided, I used a Random Forest Classifier as one possible choice for classifying planets based on the "koi\_disposition" column. The selection of this algorithm was not based on an exhaustive analysis of alternatives but rather on the algorithm's general suitability for many classification tasks. Here's a brief explanation:

While Random Forest is a robust and versatile algorithm, the choice of algorithm should also depend on the characteristics of the data and the specific problem. Random Forest Classifier contains: Versatility, Ensemble Method, Robustness, Feature Importance, Good Generalization etc.

2.)What are the different tuning methods used for the algorithm?

Tuning for the Random Forest algorithm is essential to optimize its performance. There are several methods :

>Grid Search : For example: "sci kit-learn"

>Random Search

>Manual Tuning

The choice of the tuning method depends on the available computational resources, the size of the search space, and the specific needs of the project. Grid search and random search are common starting points for tuning.

3.)Did you consider any other choice of algorithm?Why or why not?

As the choice of Random Forest Classifier algorithm was suitable with the specific characteristics of the data, the problem's requirements and my expertise with the chosen algorithm. Algorithms like Deep learning methods, such as neural networks, can be powerful for complex classification problems. And Support Vector Machines (SVM) is effective for both linear and non-linear classification problems. It's often a good practice to experiment with multiple algorithms and compare their performance using appropriate evaluation metrics to determine which one works best for the problem. The choice of this particular algorithm depends on various factors, and it is important to experiment and compare different algorithms to find the one that performs best for the given datasets.

4.) What is the accuracy?

The accuracy is a measure of the proportion of correctly classified instances out of the total instances in the test set. The relevant code which was generated in the problem:

accuracy = accuracy\_score(y\_test, predictions)

print(f"Accuracy: {accuracy}")

The accuracy value is stored in the accuracy variable, and it's printed to the console. The accuracy is expressed as a decimal value between 0 and 1, where 1 represents 100% accuracy (perfect classification) and 0 represents 0% accuracy (no correct classifications).

5.) What are the different types of metrics that can be used to evaluate the model?

There are several metrics commonly used to evaluate the performance of classification models. The choice of which metrics to use depends on the specific characteristics of the data and the goals of the model.

Accuracy, Precision, Recall (Sensitivity or True Positive Rate), Confusion metrics, Specificity (True Negative Rate)etc.

To evaluate the model some of these types of metrics is used on basis of the specific goals and constraints of the project.