Project Description: Wine Classification with scikit-learn

Overview

In this project, you will work on the Wine dataset to classify types of wine using the scikit-learn library in Python. You will apply four different machine learning algorithms:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. K-Nearest Neighbors (KNN)

You are required to deliver two versions of your solution:

- **Function-wise Version**: Implement the solution using functions.
- **OOP-wise Version**: Implement the solution using Object-Oriented Programming (OOP) principles.

Objectives

Your project should cover the following topics:

1. **Data Loading and Quality Check**

- Load the Wine dataset from scikit-learn.

2. **Data Profiling**

- Create a data profile report which includes statistics of the features, such as mean, median, range, and standard deviation.
- use the Data Profiler to do this task

3. **Data Preprocessing**

- Handle missing values if any.
- Perform feature scaling or normalization if necessary.
- Split the dataset into training and testing sets.

4. **Model Training with Cross-Validation**

- Use cross-validation techniques to train your models and ensure that your model doesn't overfit.
- Evaluate model performance using appropriate metrics (e.g., accuracy).

5. **Hyperparameter Tuning with GridSearch**

- Utilize GridSearchCV for hyperparameter tuning to improve model performance.

6. **Logging**

- Implement logging throughout the project to log important steps of the data processing and model training phases.

7. **Configuration Management**

- Use a `config.py` file along with a `config.json` to manage major parameters (e.g., model parameters, file paths).
- Think about what parameters you might need to store and manage them effectively.

8. **Results Storage**

- Store the results from different training phases into an SQLite database. Decide on the metrics and information you want to store.

9. **Model Persistence**

- Save the trained models using both pickle and Joblib for later use or deployment.

10. **Project Organization**

- When the training cycle starts, create a folder with the current date and time to store all your results (data profiles, models, SQLite database, etc.).
- For each new training cycle, a new folder should be created to keep the results organized and accessible for future reference.

Deliverables

- Source code for both the function-wise and OOP-wise versions.
- A report detailing your findings, including data profiling, model performance, and insights on the dataset and models used.
- SQLite database file containing the results of the training phases.
- Pickle and Joblib files of the trained models.
- Configuration files used in the project.
- Log files capturing the important steps and outcomes of the project.

Deadline: Monday 18.03.2024 15:00