

Nhập môn học máy

Kiểm tra cuối kỳ

Trong bài kiểm tra này, các bạn sẽ thực hiện trên tập dữ liệu Mobile Price Classification được lấy từ [Kaggle](#).

Tập dữ liệu cho biết một số đặc trưng của các dòng điện thoại và các khoảng giá bán tương ứng được mô tả ở bảng dưới đây:

No	name	Type	Description
1	id	text	ID
2	battery_power	integer	Total energy a battery can store in one time measured in mAh
3	blue	integer	Has bluetooth or not
4	clock_speed	float	speed at which microprocessor executes instructions
5	dual_sim	integer	Has dual sim support or not
6	fc	integer	Front Camera mega pixels
7	four_g	integer	Has 4G or not
8	int_memory	integer	Internal Memory in Gigabytes
9	m_dep	float	Mobile Depth in cm
10	mobile_wt	integer	Weight of mobile phone
11	n_cores	integer	Number of cores of processor
12	pc	integer	Primary Camera mega pixels
13	px_height	integer	Pixel Resolution Height
14	px_width	integer	Pixel Resolution Width
15	ram	integer	Random Access Memory in Megabytes
16	sc_h	integer	Screen Height of mobile in cm
17	sc_w	integer	Screen Width of mobile in cm
18	talk_time	integer	longest time that a single battery charge will last when you are
19	three_g	integer	Has 3G or not
20	touch_screen	integer	Has touch screen or not
21	wifi	integer	Has wifi or not
22	price_range	integer	This is the target variable with value of 0(low cost); 1(medium cost); 2(high cost) and 3(very high cost).

Nhiệm vụ chính của bài toán này là dựa trên đặc trưng của các dòng điện thoại có thể dự đoán được khoảng giá bán (**price_range**) của dòng điện thoại đó.

Tập dữ liệu gồm 3000 dòng được chia thành 2 tập: tập huấn luyện gồm 2000 dòng và tập đánh giá gồm 1000 dòng.

***Yêu cầu 1. Chọn 2 trong các mô hình học máy đã biết để giải quyết bài toán trên. Sử dụng file **train.csv** để huấn luyện và file **test.csv** để kiểm tra.**

- **Lưu ý:**
 - Thực hiện yêu cầu trên Jupyter Notebook (*.ipynb)
 - Các mô hình được cài đặt thủ công sử dụng các thư viện tính toán và trực quan hóa dữ liệu như **numpy**, **pandas**, **matplotlib**, Không sử dụng các mô hình được cài đặt sẵn.
 - Khuyến khích khảo sát hiệu suất mô hình bằng cách điều chỉnh các siêu tham số (nếu có), hoặc thử nghiệm với các cách tiền xử lý dữ liệu khác nhau.
 - Trọng số sau khi huấn luyện mô hình (nếu có) sẽ lưu vào **weights.json**
 - Các mô hình xây dựng dựa trên template sau (các hàm bổ sung được viết trong class này, nhưng đảm bảo phải có các hàm bên dưới):

```
class ModelTrainer:
    def __init__(self, **kwargs):
        """
        Initialize the ModelTrainer

        Parameters:
        - **kwargs: Additional keyword arguments for the model initialization.
        """
        pass

    def load_model_weights(self, filepath):
        """
        Load pre-trained weights for the model.

        Parameters:
        - filepath (str): Path to the file containing model weights.
        """
        pass

    def train(self, X_train, y_train):
        """
        Train the model on the given training data.

        Parameters:
        - X_train (array-like): Feature matrix for training.
        - y_train (array-like): Target vector for training.
        """
        pass

    def predict(self, X):
        """
        Predict using the trained model.
```

```

        Parameters:
        - X (array-like): Feature matrix for prediction.

        Returns:
        - Predictions (array-like): Model predictions.
        """
        pass

    def evaluate(self, X_test, y_test):
        """
        Evaluate the model on the test data.

        Parameters:
        - X_test (array-like): Feature matrix for testing.
        - y_test (array-like): True target values for testing.

        Returns:
        - Metric (float): Accuracy for classifiers,
        """
        pass

```

2. Thực hiện viết báo cáo để nộp. Nội dung báo cáo gồm:

2.1. Giới thiệu 2 mô hình học máy (đã sử dụng để giải quyết nhiệm vụ trên), trả lời các câu hỏi sau:

- Bài toán mà mô hình quan tâm.
- Vì sao chọn mô hình để giải quyết nhiệm vụ này?
- Cách thức hoạt động của mô hình (cách mô hình tối ưu tham số, các siêu tham số,). Viết ngắn gọn.
- Ưu và nhược điểm của mô hình
- Các giải quyết các nhược điểm của mô hình

2.2. Báo cáo kết quả thực nghiệm trên từng mô hình, nêu rõ các chủ điểm sau:

- Mô tả ngắn gọn quá trình tiền xử lý dữ liệu trước khi đưa vào mô hình (nếu có)
- Mô tả ngắn gọn quá trình huấn luyện
- Độ lỗi trong quá trình huấn luyện
- Độ lỗi trên tập đánh giá
- Các khám phá khác (nếu có)

3. Bài nộp là được zip lại và đặt tên là [MSSV].zip chứa các file:

- Báo cáo: [MSSV].pdf
- File code: [MSSV].ipynb
- Trọng số mô hình (nếu có): weights.json

4. Các bài giống nhau sẽ bị đánh giá 0 điểm cuối kỳ

Chúc các bạn hoàn thành tốt!
