

▼ HW3

B0928022-杜云驊

Colab-link: <https://colab.research.google.com/drive/1cfF-B5GfiacrxDcYcYLQQ0Cke3MPXIUq?usp=sharing>

```
1 class Movie:
2     def __init__(self, name, intro, label) -> None:
3         self.name = name
4         self.intro = intro
5         self.label = label

1 import json
2 path = "hw3.json"
3 movie_data = []
4 movies = []
5
6 with open(path) as f:
7     movie_data = json.load(f)
8
9 for movie in movie_data:
10     try:
11         label = movie['label'][0]
12     except:
13         label = 'NA'
14     m = Movie(movie['cname'], movie['intro'], label)
15     movies.append(m)
16
17 print(movie_data[0])
18 print(movies[0].label)

{'doc_id': 1, 'cname': '一世狂野', 'ename': 'Blow', 'pagerank': 1.1489996653.
劇情
```

```
1 with open("inverted.json") as f:
2     inverted_index = json.load(f)
3
4 key_map = {}
5 x = 0
6
7 for key in inverted_index.keys():
8     if not key in key_map.keys():
9         key_map[key] = x
10        x += 1
```



```
1 from sklearn.model_selection import train_test_split
2 from sklearn.neighbors import KNeighborsClassifier
3
4 dataframe.columns = dataframe.columns.astype(str)
5
6 X_data = dataframe.drop(columns=["name", "label"])
7 y_data = dataframe["label"]
8
9 X_data
10
11 X_train, X_test, y_train, y_test = train_test_split(X_data, y_data,
12 clf = KNeighborsClassifier()
13 clf.fit(X_train,y_train)
14 clf.score(X_test,y_test)

0.3592233009708738
```

```
1 from sklearn import tree
2
3 clf = tree.DecisionTreeClassifier()
4 clf.fit(X_train,y_train)
5 clf.score(X_test,y_test)

0.34951456310679613
```