Perbandingan LabelEncoder dan OneHotEncoder Dalam penerapan Machine Learning

ACHMAD RERO SINDUNATA

DATASET

LINK DOWNLOAD :
https://www.kaggle.com/datasets/ibrahimbahbah/drug200

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Age          200 non-null    int64  
 1   Sex          200 non-null    object 
 2   BP           200 non-null    object 
 3   Cholesterol  200 non-null    object 
 4   Na_to_K      200 non-null    float64
 5   Drug         200 non-null    object 
dtypes: float64(1), int64(1), object(4)
memory usage: 9.5+ KB
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	DrugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	DrugY

DECISION TREE

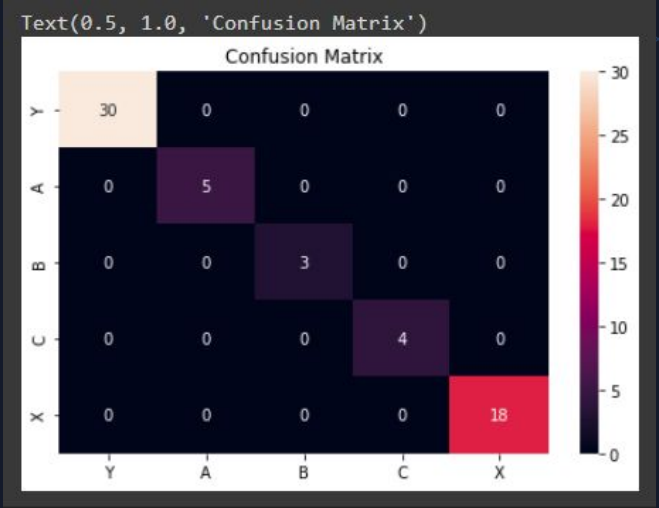
```
Accuracy: 1.00

Micro Precision: 1.00
Micro Recall: 1.00
Micro F1-score: 1.00

Macro Precision: 1.00
Macro Recall: 1.00
Macro F1-score: 1.00

Weighted Precision: 1.00
Weighted Recall: 1.00
Weighted F1-score: 1.00

Classification Report

              precision    recall  f1-score   support

     Class 0       1.00      1.00      1.00        30
     Class 1       1.00      1.00      1.00         5
     Class 2       1.00      1.00      1.00         3
     Class 3       1.00      1.00      1.00         4
     Class 4       1.00      1.00      1.00        18

    accuracy                           1.00        60
   macro avg       1.00      1.00      1.00        60
weighted avg       1.00      1.00      1.00        60
```

LabelEncoder

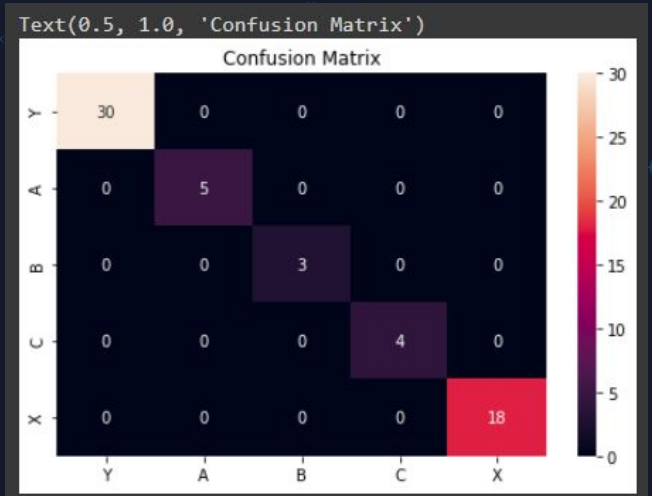
```
Accuracy: 1.00

Micro Precision: 1.00
Micro Recall: 1.00
Micro F1-score: 1.00

Macro Precision: 1.00
Macro Recall: 1.00
Macro F1-score: 1.00

Weighted Precision: 1.00
Weighted Recall: 1.00
Weighted F1-score: 1.00

Classification Report

              precision    recall  f1-score   support

     Class 0       1.00      1.00      1.00        30
     Class 1       1.00      1.00      1.00         5
     Class 2       1.00      1.00      1.00         3
     Class 3       1.00      1.00      1.00         4
     Class 4       1.00      1.00      1.00        18

    accuracy                           1.00        60
   macro avg       1.00      1.00      1.00        60
weighted avg       1.00      1.00      1.00        60
```

OneHotEncoder

DECISION TREE LANJUT



LabelEncoder



OneHotEncoder

Percobaan pertama menggunakan metode Decision Tree (Default). Dilihat dari percobaan ini, antara menggunakan LabelEncoder dan OneHotEncoder tidak terjadi perbedaan.

SUPPORT VECTOR MACHINE (SVM)

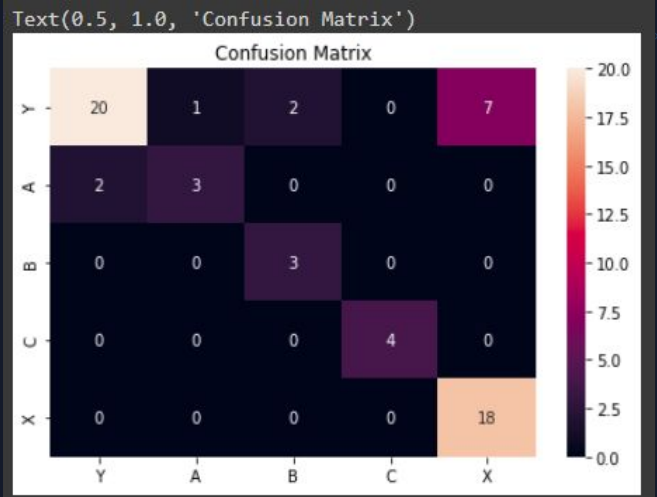
```
Accuracy: 0.80

Micro Precision: 0.80
Micro Recall: 0.80
Micro F1-score: 0.80

Macro Precision: 0.80
Macro Recall: 0.85
Macro F1-score: 0.80

Weighted Precision: 0.83
Weighted Recall: 0.80
Weighted F1-score: 0.80

Classification Report

              precision    recall  f1-score   support

     Class 0       0.91      0.67      0.77        30
     Class 1       0.75      0.60      0.67         5
     Class 2       0.60      1.00      0.75         3
     Class 3       1.00      1.00      1.00         4
     Class 4       0.72      1.00      0.84        18

    accuracy                           0.80        60
   macro avg       0.80      0.85      0.80        60
weighted avg       0.83      0.80      0.80        60
```

LabelEncoder

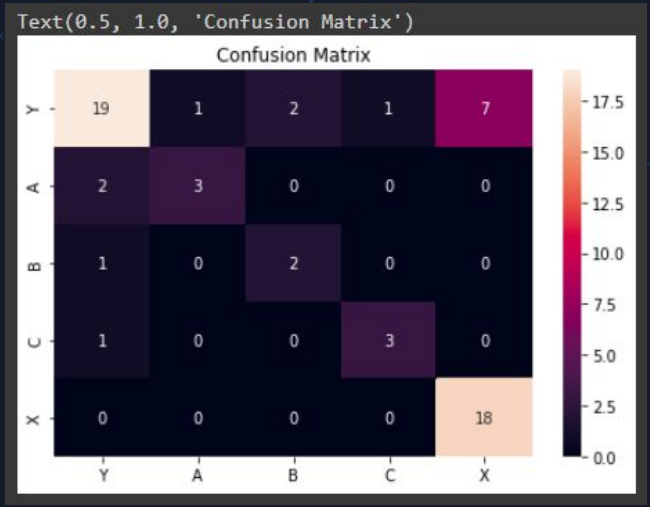
```
Accuracy: 0.75

Micro Precision: 0.75
Micro Recall: 0.75
Micro F1-score: 0.75

Macro Precision: 0.71
Macro Recall: 0.73
Macro F1-score: 0.71

Weighted Precision: 0.77
Weighted Recall: 0.75
Weighted F1-score: 0.74

Classification Report

              precision    recall  f1-score   support

     Class 0       0.83      0.63      0.72        30
     Class 1       0.75      0.60      0.67         5
     Class 2       0.50      0.67      0.57         3
     Class 3       0.75      0.75      0.75         4
     Class 4       0.72      1.00      0.84        18

    accuracy                           0.75        60
   macro avg       0.71      0.73      0.71        60
weighted avg       0.77      0.75      0.74        60
```

OneHotEncoder

SUPPORT VECTOR MACHINE (SVM) LANJUT



LabelEncoder



OneHotEncoder

Percobaan kedua menggunakan metode Support Vector Machine (SVM) (Default). Dilihat dari percobaan ini, antara menggunakan LabelEncoder dan OneHotEncoder terdapat perbedaan. Saat menggunakan LabelEncoder performa SVM lebih baik daripada menggunakan OneHotEncoder.

K-NEAREST NEIGHBOR (KNN)

Accuracy: 0.73

Micro Precision: 0.73
Micro Recall: 0.73
Micro F1-score: 0.73

Macro Precision: 0.72
Macro Recall: 0.76
Macro F1-score: 0.73

Weighted Precision: 0.75
Weighted Recall: 0.73
Weighted F1-score: 0.73

Classification Report

	precision	recall	f1-score	support
Class 0	0.82	0.60	0.69	30
Class 1	0.60	0.60	0.60	5
Class 2	0.50	0.67	0.57	3
Class 3	1.00	1.00	1.00	4
Class 4	0.68	0.94	0.79	18
accuracy			0.73	60
macro avg	0.72	0.76	0.73	60
weighted avg	0.75	0.73	0.73	60

LabelEncoder

Accuracy: 0.73

Micro Precision: 0.73
Micro Recall: 0.73
Micro F1-score: 0.73

Macro Precision: 0.72
Macro Recall: 0.76
Macro F1-score: 0.73

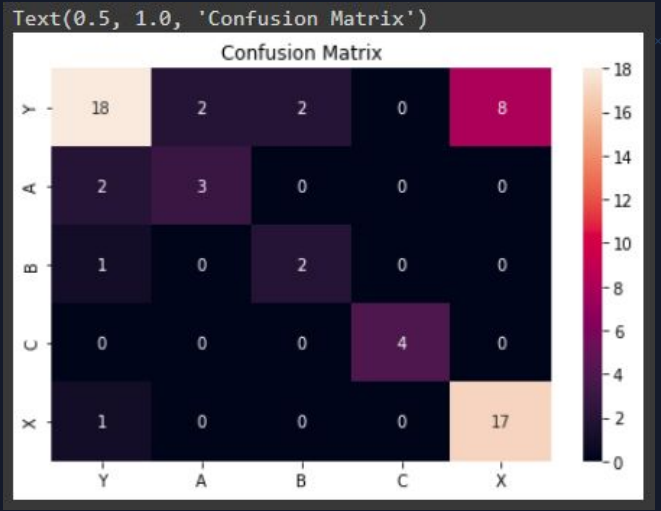
Weighted Precision: 0.75
Weighted Recall: 0.73
Weighted F1-score: 0.73

Classification Report

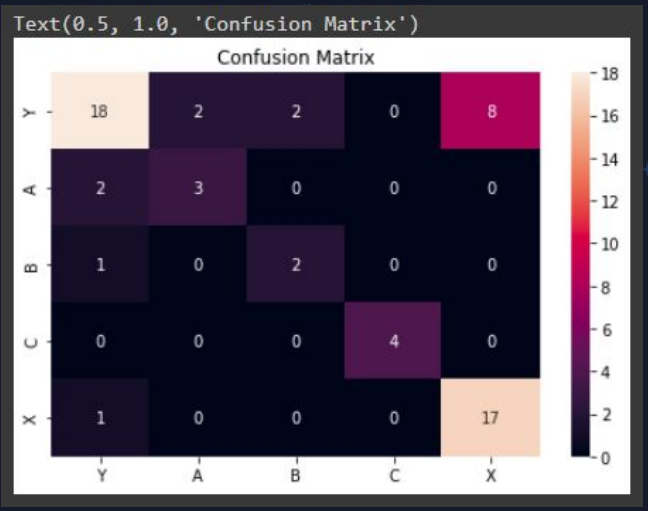
	precision	recall	f1-score	support
Class 0	0.82	0.60	0.69	30
Class 1	0.60	0.60	0.60	5
Class 2	0.50	0.67	0.57	3
Class 3	1.00	1.00	1.00	4
Class 4	0.68	0.94	0.79	18
accuracy			0.73	60
macro avg	0.72	0.76	0.73	60
weighted avg	0.75	0.73	0.73	60

OneHotEncoder

K-NEAREST NEIGHBOR (KNN)



LabelEncoder



OneHotEncoder

Percobaan ketiga menggunakan metode K-Nearest Neighbor (n = 3). Dilihat dari percobaan ini, antara menggunakan LabelEncoder dan OneHotEncoder tidak terjadi perbedaan.

KESIMPULAN

1. Menggunakan Decision Tree dalam percobaan ini tidak terjadi perbedaan.
2. Menggunakan SVM dalam percobaan ini terdapat perbedaan.
3. Menggunakan KNN dalam percobaan ini tidak terjadi perbedaan.

LINK GITHUB

https://github.com/rerosindunata/LabelEncoder-vs-OneHoteEncoder


TERIMA KASIH