

Implementasi Metode Decision Tree untuk Penyakit Hepatitis

Nama : Irvan Akbar Febriansyah

NIM : 19090099

Kelas : 4A

1. PENDAHULUAN

Penyakit Hepatitis adalah penyakit yang disebabkan oleh beberapa jenis virus yang menyerang dan menyebabkan peradangan serta merusak sel-sel organ hati manusia. Hepatitis merupakan salah satu penyakit yang membahayakan jika tidak segera ditangani. Penyakit yang menyerang hati atau liver ini semakin berbahaya karena gejalanya yang tidak selalu tampak. Fungsi utama dari hati atau liver adalah menyaring racun-racun yang ada pada darah. Selain itu, masih ada sekitar 500 fungsi lain dari hati.

Jika seseorang menderita hepatitis, yang merupakan peradangan pada hati atau liver ini, dapat menghancurkan kesehatan orang tersebut secara keseluruhan karena racun tetap mengendap pada darah dan merusak atau mengganggu kerja organ lain. Akibat lainnya adalah hati menolak darah yang mengalir sehingga tekanan darah menjadi tinggi dan pecahnya pembuluh darah.

Hepatitis dikategorikan dalam beberapa golongan, diantaranya hepatitis A, B, C, D, E. Penderita penyakit Hepatitis umumnya cenderung lebih banyak mengalami golongan hepatitis B dan hepatitis C. Penyakit ini memiliki penyebab antara lain, virus, komplikasi dari penyakit lain, alkohol, obat-obatan atau zat kimia sampai karena penyakit autoimun (Sari W dan Indrawati L, 2008:16).

Penelitian terkait hepatitis sudah pernah dilakukan sebelumnya, yaitu “Sistem Diagnosa Penyakit Hepatitis dengan menggunakan Metode Naïve Bayesian” oleh EstuKarunianingtyas. Penelitian tersebut menggunakan data mining untuk menentukan pasien tertentu masuk pada kelas hepatitis A, B atau C berdasarkan gejala. Akurasi pada penelitian sebelumnya tergolong rendah karena hanya mendapatkan 51,11% untuk kombinasi gejala dan 44,44% untuk per gejala.

Jadi, pada penelitian kali ini akan memprediksi akurasi seseorang apakah terindikasi terkena penyakit hepatitis atau tidak berdasarkan variabel umur, jenis kelamin, serta berbagai gejala-gejala lainnya seperti fatigue yaitu kondisi di mana tubuh selalu merasa lelah, lesu, atau kurang tenaga. Kemudian gejala lain yaitu malaise dimana tubuh merasa lemas, tidak nyaman dan badan merasa kurang fit.

2. METODE PENELITIAN

Decision tree akan memperlihatkan factor-faktor kemungkinan (probabilitas) yang akan mempengaruhi alternatif –alternatif dalam penentuan hepatitis, disertai dengan prediksi hasil akhir untuk penentuan dengan menggunakan Decision Tree. Data dalam Decision Tree dinyatakan dalam bentuk table dengan atribut dan record. Atribut menyatakan parameter yang dibuat sebagai kriteria dalam pembentukan tree.

Decision Tree adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (leaf) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah Decision Tree adalah node akar (root) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya Decision Tree melakukan strategi pencarian secara top-down untuk solusinya.

Pada penelitian kali ini yaitu memprediksi akurasi dengan metode decision tree orang mengidap hepatitis atau tidaknya berdasarkan variabel dependen sebagai variabel target yaitu Suspect (Terkena) atau Not Suspect (Tidak Terkena) dan variabel Age (Usia), Gender (Jenis Kelamin), Fatigue (Rasa Kelelahan) , Malaise (Rasa Lemas) , LiverFirm, LiverBig dan Histology.

3. HASIL DAN PEMBAHASAN

1. Mengimport Library

```
#import library
import pandas as pd
import pydotplus
from IPython.display import Image
from sklearn import tree
```

2. Mengimport Dataset

import Dataset

```
[ ] data = pd.read_csv('hepatitis1.csv')
data.head()
```

	ID	target	age	gender	fatigue	malaise	anorexia	liverBig	liverFirm	spleen	spiders	ascites	varices	bili	alk	sgot	albu	protine	histology
0	1	Suspect	30	2	2	2	2	1	2	2	2	2	2	1	85	18	4	?	1
1	2	Suspect	50	1	1	2	2	1	2	2	2	2	2	0.9	135	42	3.5	?	1
2	3	Suspect	78	1	1	2	2	2	2	2	2	2	2	0.7	96	32	4	?	1
3	4	Suspect	31	1	2	2	2	2	2	2	2	2	2	0.7	46	52	4	80	1
4	5	Suspect	34	1	2	2	2	2	2	2	2	2	2	1	?	200	4	?	1

3. Preprocessing data untuk memperbaiki data yang kosong atau noise

Preprocessing data untuk memperbaiki data yang kosong atau noise

```
▶ from sklearn.preprocessing import LabelEncoder
enc = LabelEncoder()
data['target']=enc.fit_transform(data['target'].values)
data['age']=enc.fit_transform(data['age'].values)
data['gender']=enc.fit_transform(data['gender'].values)
data['fatigue']=enc.fit_transform(data['fatigue'].values)
data['malaise']=enc.fit_transform(data['malaise'].values)
data['anorexia']=enc.fit_transform(data['anorexia'].values)
data['liverFirm']=enc.fit_transform(data['liverFirm'].values)
data['spleen']=enc.fit_transform(data['spleen'].values)
data['spiders']=enc.fit_transform(data['spiders'].values)
data['ascites']=enc.fit_transform(data['ascites'].values)
data['bili']=enc.fit_transform(data['bili'].values)
data['varices']=enc.fit_transform(data['varices'].values)
data['histology']=enc.fit_transform(data['histology'].values)
data['liverBig']=enc.fit_transform(data['liverBig'].values)
data['albu']=enc.fit_transform(data['albu'].values)
data['alk']=enc.fit_transform(data['alk'].values)
data['protime']=enc.fit_transform(data['protime'].values)
data.head()
```

4. Data yang sudah di preprocessing

	ID	target	age	gender	fatigue	malaise	anorexia	liverBig	liverFirm	spleen	spiders	ascites	varices	bili	alk	sgot	albu	protime	histology
0	1	1	9	1	1	1	1	0	1	1	1	1	1	7	76	18	16	44	0
1	2	1	29	0	0	1	1	0	1	1	1	1	1	6	17	42	11	44	0
2	3	1	48	0	0	1	1	1	1	1	1	1	1	4	82	32	16	44	0
3	4	1	10	0	1	1	1	1	1	1	1	1	1	4	49	52	16	40	0
4	5	1	13	0	1	1	1	1	1	1	1	1	1	7	83	200	16	44	0

Semua data yang ada di dataset saya preprocessing agar data data yang kosong, rusak atau noise dapat diperbaiki

5. Split Data Dependent dan independent

Proses Split data dependent dan independent

saya menyisihkan / drop atribut yang tidak saya gunakan, dan menyisakan variabel independent di atr_data

```
atr_data = data.drop(columns='target')
atr_data = atr_data.drop(columns='ID')
atr_data = atr_data.drop(columns='gender')
atr_data = atr_data.drop(columns='anorexia')
atr_data = atr_data.drop(columns='spleen')
atr_data = atr_data.drop(columns='spiders')
atr_data = atr_data.drop(columns='ascites')
atr_data = atr_data.drop(columns='varices')
atr_data = atr_data.drop(columns='bili')
atr_data = atr_data.drop(columns='alk')
atr_data = atr_data.drop(columns='sgot')
atr_data = atr_data.drop(columns='albu')
atr_data = atr_data.drop(columns='protime')
atr_data.head()
```

	age	fatigue	malaise	liverBig	liverFirm	histology
0	9	1	1	0	1	0
1	29	0	1	0	1	0
2	48	0	1	1	1	0
3	10	1	1	1	1	0
4	13	1	1	1	1	0

Pertama saya drop data yang tidak saya butuh kan, dan Untuk Variabel Independent saya menggunakan age, fatigue, malaise, liverbig, liverFiem dan histology dan saya simpan di variable atr_data.

6. Variabel Dependent

```
[ ] cls_data=data['target']
    cls_data.head()

0    1
1    1
2    1
3    1
4    1
Name: target, dtype: int64
```

Variabel dependent saya menggunakan target sebagai hasil hepatitis atau tidaknya

7. Mengecek ulang data dependent dan independent

```
[ ] atr_data = atr_data.dropna()
```

```
▶ cls_data = cls_data.dropna()
```

8. import Library decision tree untuk pembuatan model

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.tree import DecisionTreeClassifier
```

9. pembuatan model testing decision tree setelah split data test dan train

```
x_train, x_test, y_train, y_test = train_test_split(atr_data, cls_data, test_size=0.3, random_state=0)
tree_data = DecisionTreeClassifier(random_state=0)
tree_data.fit(x_train, y_train)
```

```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',
                        max_depth=None, max_features=None, max_leaf_nodes=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, presort='deprecated',
                        random_state=0, splitter='best')
```

Untuk model testing decision tree untuk data training saya menggunakan 30% data dan sisanya untuk testing

10. Menampilkan nilai akurasi dari model

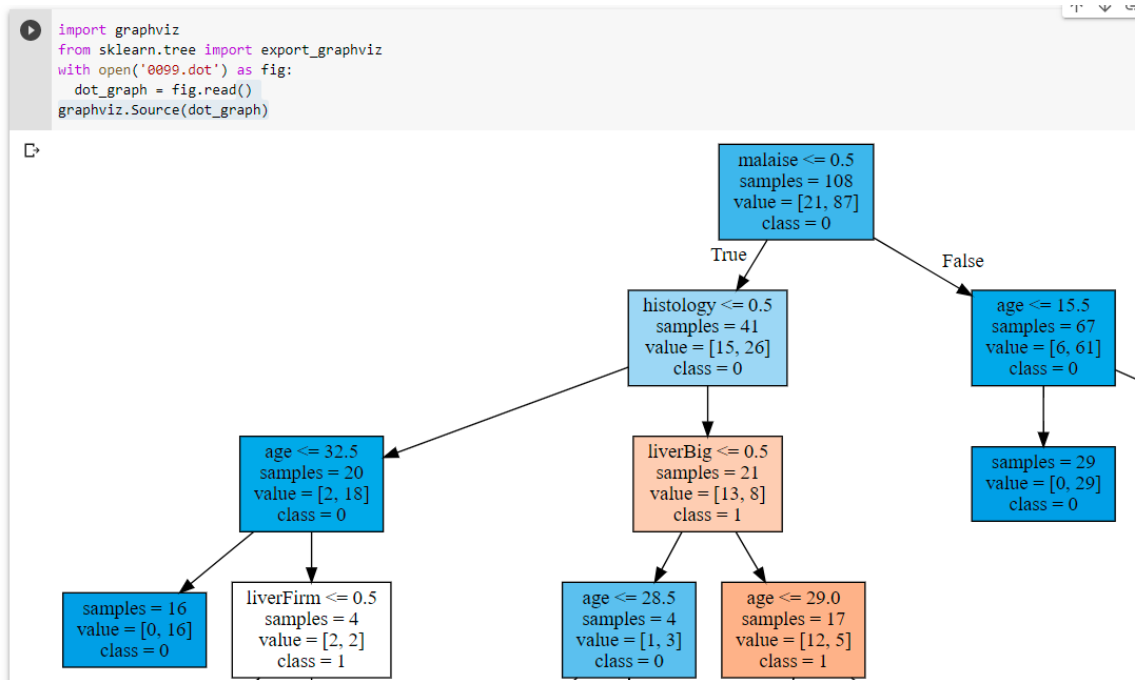
```
print ('nilai akurasi:', tree_data.score(x_test, y_test))
```

```
nilai akurasi: 0.6595744680851063
```

11. Mengimport graphviz untuk memvisualisasikan data

```
from sklearn.tree import export_graphviz
export_graphviz(tree_data, out_file='0099.dot', class_names=["1","0"],
                feature_names=x.columns, impurity=False, filled=True)
```

12. Tampilan data setelah di coba menggunakan decision tree



DAFTAR PUSTAKA

- Asrini, Christina Wienda. (2013). Identifikasi Penyakit Hepatitis dengan Pendekatan *Agglomerative Hierarchical Clustering*. Skripsi Fakultas Sains dan Teknologi, Jurusan Teknik Informatika, Universitas Sanata Dharma, Yogyakarta.
- Rusdah. 2010. Pengembangan Decision Support System Untuk Mendukung Analisis Pengambilan Keputusan. *Telematika Mkom* , 32-40.
-