diabetes.ipynb

File Edit View Insert Runtime Tools Help Last edited on Aug 25, 2022

Comment

+ Code + Text

```
1 import pandas as pd
2 import seaborn as sb
3 import numpy as np
4 import matplotlib.pyplot as plt
```

```
1 dbts = pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Diabetes.csv')
```

```
1 dbts.head()
```

	pregnancies	glucose	diastolic	triceps	insulin	bmi	dpf	age	diabetes
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
1 dbts.info()
```

```
<class 'pandas.core.frame.DataF
RangeIndex: 768 entries, 0 to 7
Data columns (total 9 columns):
 #   Column       Non-Null Coun
---  ------       -------------
 0   pregnancies  768 non-null
 1   glucose      768 non-null
 2   diastolic    768 non-null
 3   triceps      768 non-null
 4   insulin      768 non-null
 5   bmi          768 non-null
 6   dpf          768 non-null
 7   age          768 non-null
 8   diabetes     768 non-null
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
1 dbts.isnull()
```

	pregnancies	glucose	diastolic	triceps	insulin	bmi	dpf	age	diabetes
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
763	False	False	False	False	False	False	False	False	False
764	False	False	False	False	False	False	False	False	False
765	False	False	False	False	False	False	False	False	False
766	False	False	False	False	False	False	False	False	False
767	False	False	False	False	False	False	False	False	False

768 rows × 9 columns

```
1 dbts.nunique()
```

```
pregnancies     17
glucose        136
diastolic       47
triceps         51
insulin        186
bmi            248
dpf            517
age             52
diabetes         2
dtype: int64
```

```
1 dbts.columns
```

```
Index(['pregnancies', 'glucose'
'diastolic', 'triceps', 'insuli
'bmi',
       'dpf', 'age', 'diabetes'
      dtype='object')
```

```
1 y = dbts['diabetes']
```

```
1 x = dbts[['pregnancies', 'glucose', 'diastolic', 'triceps', 'insulin', 'bmi',
2       'dpf', 'age']]
```

```
1 #from sklearn.linear_model import LogisticRegression
```

```
1 #model = LogisticRegression(max_iter=200)
```

```
1 from sklearn.ensemble import RandomForestClassifier
2 model = RandomForestClassifier()
```

```
1 model.fit(x, y)
```

```
RandomForestClassifier()
```

```
1 y_pred = model.predict(x)
```

```
1 from sklearn.metrics import accuracy_score
```

```
1 accuracy_score(y, y_pred)
```