```
#import library
import pandas as pd


#import data
disease = pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/MultipleDiseaseP


#view data
disease.head()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shiv
0	1	1	1	0	
1	0	1	1	0	
2	1	0	1	0	
3	1	1	0	0	
4	1	1	1	0	

5 rows × 133 columns

```
#info of data
disease.info(verbose=True)

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 4920 entries, 0 to 4919
    Data columns (total 133 columns):
     #    Column                    Dtype
    ---   ------                    -----
     0    itching                   int64
     1    skin_rash                 int64
     2    nodal_skin_eruptions      int64
     3    continuous_sneezing       int64
     4    shivering                 int64
     5    chills                    int64
     6    joint_pain                int64
     7    stomach_pain              int64
     8    acidity                   int64
     9    ulcers_on_tongue          int64
     10   muscle_wasting            int64
     11   vomiting                  int64
     12   burning_micturition       int64
```

13	spotting_ urination	int64
14	fatigue	int64
15	weight_gain	int64
16	anxiety	int64
17	cold_hands_and_feets	int64
18	mood_swings	int64
19	weight_loss	int64
20	restlessness	int64
21	lethargy	int64
22	patches_in_throat	int64
23	irregular_sugar_level	int64
24	cough	int64
25	high_fever	int64
26	sunken_eyes	int64
27	breathlessness	int64
28	sweating	int64
29	dehydration	int64
30	indigestion	int64
31	headache	int64
32	yellowish_skin	int64
33	dark_urine	int64
34	nausea	int64
35	loss_of_appetite	int64
36	pain_behind_the_eyes	int64
37	back_pain	int64
38	constipation	int64
39	abdominal_pain	int64
40	diarrhoea	int64
41	mild_fever	int64
42	yellow_urine	int64
43	yellowing_of_eyes	int64
44	acute_liver_failure	int64
45	fluid_overload	int64
46	swelling_of_stomach	int64
47	swelled_lymph_nodes	int64
48	malaise	int64
49	blurred_and_distorted_vision	int64
50	phlegm	int64
51	throat_irritation	int64
52	redness_of_eyes	int64

```
#summary statistics
disease.describe()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing
count	4920.000000	4920.000000	4920.000000	4920.000000
mean	0.137805	0.159756	0.021951	0.045122
std	0.344730	0.366417	0.146539	0.207593
min	0.000000	0.000000	0.000000	0.000000

```
#check for missing value
disease.isnull().sum()
```

```
itching                 0
skin_rash               0
nodal_skin_eruptions    0
continuous_sneezing     0
shivering               0
                       ..
inflammatory_nails      0
blister                 0
red_sore_around_nose    0
yellow_crust_ooze       0
prognosis               0
Length: 133, dtype: int64
```

```
#check for categories
disease.prognosis.nunique()
```

```
41
```

```
disease.nunique()
```

```
itching                 2
skin_rash               2
nodal_skin_eruptions    2
continuous_sneezing     2
shivering               2
                       ..
inflammatory_nails      2
blister                 2
red_sore_around_nose    2
yellow_crust_ooze       2
prognosis              41
Length: 133, dtype: int64
```

```
#correlation
disease.corr()
```

	itching	skin_rash	nodal_skin_eruptions	continuous
itching	1.000000	0.318158	0.326439	
skin_rash	0.318158	1.000000	0.298143	
nodal_skin_eruptions	0.326439	0.298143	1.000000	
continuous_sneezing	-0.086906	-0.094786	-0.032566	
shivering	-0.059893	-0.065324	-0.022444	
...	...	...	...	
small_dents_in_nails	-0.061573	0.331087	-0.023073	
inflammatory_nails	-0.061573	0.331087	-0.023073	
blister	-0.061573	0.331087	-0.023073	
red_sore_around_nose	-0.061573	0.331087	-0.023073	
yellow_crust_ooze	-0.061573	0.331087	-0.023073	

```python
#visualize pairplot
import seaborn as sns


#sns.pairplot(disease)
import matplotlib.pyplot as plt


disease['prognosis'].value_counts()
```

```
Fungal infection                          120
Hepatitis C                               120
Hepatitis E                               120
Alcoholic hepatitis                       120
Tuberculosis                              120
Common Cold                               120
Pneumonia                                 120
Dimorphic hemmorhoids(piles)              120
Heart attack                              120
Varicose veins                            120
Hypothyroidism                            120
Hyperthyroidism                           120
Hypoglycemia                              120
Osteoarthristis                           120
Arthritis                                 120
(vertigo) Paroymsal  Positional Vertigo   120
Acne                                      120
Urinary tract infection                   120
Psoriasis                                 120
Hepatitis D                               120
Hepatitis B                               120
Allergy                                   120
hepatitis A                               120
```

```
GERD                            120
Chronic cholestasis             120
Drug Reaction                   120
Peptic ulcer diseae             120
AIDS                            120
Diabetes                        120
Gastroenteritis                 120
Bronchial Asthma                120
Hypertension                    120
Migraine                        120
Cervical spondylosis            120
Paralysis (brain hemorrhage)    120
Jaundice                        120
Malaria                         120
Chicken pox                     120
Dengue                          120
Typhoid                         120
Impetigo                        120
Name: prognosis, dtype: int64
```

```python
disease ['prognosis'].value_counts().plot.bar()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5365f9d310>
```
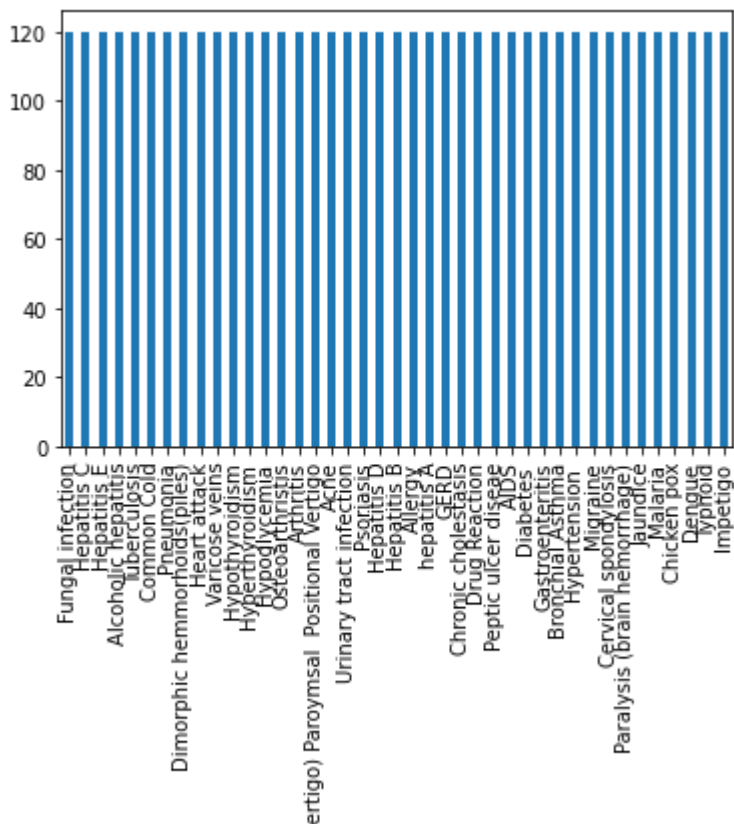


```python
#column name
disease.columns
```

```
Index(['itching', 'skin_rash', 'nodal_skin_eruptions', 'continuous_sneezing',
       'shivering', 'chills', 'joint_pain', 'stomach_pain', 'acidity',
       'ulcers_on_tongue',
       ...
```

```
            'blackheads', 'scurring', 'skin_peeling', 'silver_like_dusting',
            'small_dents_in_nails', 'inflammatory_nails', 'blister',
            'red_sore_around_nose', 'yellow_crust_ooze', 'prognosis'],
          dtype='object', length=133)
```

```
disease.shape
```

```
    (4920, 133)
```

```
#define y
y = disease['prognosis']
```

```
#define x
x = disease[['itching','skin_rash','nodal_skin_eruptions','shivering','joint_pain','stomac
```

```
#split data
from sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train, y_test = train_test_split(x,y,train_size=0.8,random_state=2529)
```

```
#verify shape
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

```
    ((3936, 15), (984, 15), (3936,), (984,))
```

```
#select model
from sklearn.ensemble import RandomForestClassifier
```

```
#train model
model.fit (x_train , y_train)
```

```
    RandomForestClassifier()
```

```
#predict with model
y_pred =model.predict(x_test)
```

```
#model accuracy
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

```
#model evoluation
accuracy_score(y_test , y_pred)
```

```
    0.2774390243902439
```

```
#model confusion matrix
confusion_matrix(y_test, y_pred)
```

```
array([[ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0, 22, ...,  0,  0,  0],
       ...,
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0]])
```

```
#model classification report
confusion_matrix(y_test, y_pred)
```

```
array([[ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0, 22, ...,  0,  0,  0],
       ...,
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0],
       [ 0,  0,  0, ...,  0,  0,  0]])
```

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	suppor
(vertigo) Paroymsal Positional Vertigo	0.00	0.00	0.00	2
AIDS	0.00	0.00	0.00	2
Acne	1.00	1.00	1.00	2
Alcoholic hepatitis	0.00	0.00	0.00	2
Allergy	1.00	0.95	0.98	2
Arthritis	0.00	0.00	0.00	2
Bronchial Asthma	0.00	0.00	0.00	2
Cervical spondylosis	0.00	0.00	0.00	2
Chicken pox	0.79	0.96	0.86	2
Chronic cholestasis	0.29	0.96	0.45	2
Common Cold	0.00	0.00	0.00	2
Dengue	1.00	0.95	0.97	2
Diabetes	0.00	0.00	0.00	2
Dimorphic hemmorhoids(piles)	0.00	0.00	0.00	2
Drug Reaction	1.00	0.90	0.95	2
Fungal infection	1.00	0.88	0.93	2
GERD	1.00	1.00	1.00	1
Gastroenteritis	0.00	0.00	0.00	2
Heart attack	0.00	0.00	0.00	2
Hepatitis B	0.00	0.00	0.00	2
Hepatitis C	0.00	0.00	0.00	2
Hepatitis D	0.00	0.00	0.00	2
Hepatitis E	0.17	0.83	0.28	1
Hypertension	0.00	0.00	0.00	2
Hyperthyroidism	0.00	0.00	0.00	2
Hypoglycemia	0.00	0.00	0.00	2
Hypothyroidism	0.03	1.00	0.06	1
Impetigo	1.00	1.00	1.00	3
Jaundice	0.00	0.00	0.00	2
Malaria	0.00	0.00	0.00	2
Migraine	1.00	0.95	0.98	2

Osteoarthristis	0.00	0.00	0.00	3
Paralysis (brain hemorrhage)	0.00	0.00	0.00	2
Peptic ulcer diseae	0.00	0.00	0.00	2
Pneumonia	0.00	0.00	0.00	2
Psoriasis	1.00	1.00	1.00	2
Tuberculosis	0.00	0.00	0.00	3
Typhoid	0.00	0.00	0.00	2
Urinary tract infection	0.00	0.00	0.00	2
Varicose veins	0.00	0.00	0.00	2
hepatitis A	0.00	0.00	0.00	2
accuracy			0.28	98
macro avg	0.25	0.30	0.26	98
weighted avg	0.24	0.28	0.24	98

```
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_classification.py:1318:
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_classification.py:1318:
  _warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_classification.py:1318:
  _warn_prf(average, modifier, msg_start, len(result))
```

```
#future prediction
x_new = x.sample()
```

```
#define x_new
x_new
```

	itching	skin_rash	nodal_skin_eruptions	shivering	joint_pain
367	0	0	0	0	0

```
#product for x_new
model.predict(x_new)
```

```
array(['Hypothyroidism'], dtype=object)
```