★ Classification Model to Identify Multiple Disease

```
# import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# import data
disease = pd.read_csv('https://github.com/ybifoundation/Dataset/raw/main/MultipleDiseasePredictio
```

```
# view data
disease.head()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chil
0	1	1	1	0	0	
1	0	1	1	0	0	
2	1	0	1	0	0	
3	1	1	0	0	0	
4	1	1	1	0	0	

5 rows × 133 columns

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

```
 76   drying_and_tingling_lips        4920 non-null   int64
 77   slurred_speech                  4920 non-null   int64
 78   knee_pain                       4920 non-null   int64
 79   hip_joint_pain                  4920 non-null   int64
 80   muscle_weakness                 4920 non-null   int64
 81   stiff_neck                      4920 non-null   int64
 82   swelling_joints                 4920 non-null   int64
 83   movement_stiffness              4920 non-null   int64
 84   spinning_movements              4920 non-null   int64
 85   loss_of_balance                 4920 non-null   int64
 86   unsteadiness                    4920 non-null   int64
 87   weakness_of_one_body_side       4920 non-null   int64
 88   loss_of_smell                   4920 non-null   int64
 89   bladder_discomfort              4920 non-null   int64
 90   foul_smell_of urine             4920 non-null   int64
 91   continuous_feel_of_urine        4920 non-null   int64
 92   passage_of_gases                4920 non-null   int64
 93   internal_itching                4920 non-null   int64
```

```
 93   internal_itching                4920 non-null   int64
 94   toxic_look_(typhos)             4920 non-null   int64
 95   depression                      4920 non-null   int64
 96   irritability                    4920 non-null   int64
 97   muscle_pain                     4920 non-null   int64
 98   altered_sensorium               4920 non-null   int64
 99   red_spots_over_body             4920 non-null   int64
 100  belly_pain                      4920 non-null   int64
 101  abnormal_menstruation           4920 non-null   int64
 102  dischromic _patches             4920 non-null   int64
 103  watering_from_eyes              4920 non-null   int64
 104  increased_appetite              4920 non-null   int64
 105  polyuria                        4920 non-null   int64
 106  family_history                  4920 non-null   int64
 107  mucoid_sputum                   4920 non-null   int64
 108  rusty_sputum                    4920 non-null   int64
 109  lack_of_concentration           4920 non-null   int64
 110  visual_disturbances             4920 non-null   int64
 111  receiving_blood_transfusion     4920 non-null   int64
 112  receiving_unsterile_injections  4920 non-null   int64
 113  coma                            4920 non-null   int64
 114  stomach_bleeding                4920 non-null   int64
 115  distention_of_abdomen           4920 non-null   int64
 116  history_of_alcohol_consumption  4920 non-null   int64
 117  fluid_overload.1                4920 non-null   int64
 118  blood_in_sputum                 4920 non-null   int64
 119  prominent_veins_on_calf         4920 non-null   int64
 120  palpitations                    4920 non-null   int64
 121  painful_walking                 4920 non-null   int64
 122  pus_filled_pimples              4920 non-null   int64
 123  blackheads                      4920 non-null   int64
 124  scurring                        4920 non-null   int64
 125  skin_peeling                    4920 non-null   int64
 126  silver_like_dusting             4920 non-null   int64
 127  small_dents_in_nails            4920 non-null   int64
 128  inflammatory_nails              4920 non-null   int64
 129  blister                         4920 non-null   int64
 130  red_sore_around_nose            4920 non-null   int64
 131  yellow_crust_ooze               4920 non-null   int64
 132  prognosis                       4920 non-null   object
dtypes: int64(132), object(1)
```

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

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shive
count	4920.000000	4920.000000	4920.000000	4920.000000	4920.00

```
# check for missing value
disease.isna().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.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_sneez
itching	1.000000	0.318158	0.326439	-0.086
skin_rash	0.318158	1.000000	0.298143	0.094
continuous_sneezing	-0.086906	-0.094786	-0.032566	1.000

```
# visualize pairplot
sns.pairplot(disease)
```

```
# column names
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)
```

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

```
# define X
X = disease.drop(['prognosis'], axis = 1)
```

```
# split data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=.30, random_state=2529)
```

```
# verify shape
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
((3444, 132), (1476, 132), (3444,), (1476,))
```

```
# select model
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
```

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

```
RandomForestClassifier()
```

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

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

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

```
1.0
```

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

```
array([[31,  0,  0, ...,  0,  0,  0],
       [ 0, 37,  0, ...,  0,  0,  0],
       [ 0,  0, 35, ...,  0,  0,  0],
       ...,
       [ 0,  0,  0, ..., 39,  0,  0],
       [ 0,  0,  0, ...,  0, 35,  0],
       [ 0,  0,  0, ...,  0,  0, 32]])
```

```
# model classification report
print(classification_report(y_test, y_pred))
```

```
                                        precision    recall  f1-score   support

(vertigo) Paroymsal  Positional Vertigo       1.00      1.00      1.00        31
                                    AIDS       1.00      1.00      1.00        37
                                    Acne       1.00      1.00      1.00        35
                    Alcoholic hepatitis       1.00      1.00      1.00        40
                                 Allergy       1.00      1.00      1.00        37
                               Arthritis       1.00      1.00      1.00        46
                          Bronchial Asthma       1.00      1.00      1.00        37
                      Cervical spondylosis       1.00      1.00      1.00        31
                               Chicken pox       1.00      1.00      1.00        29
                      Chronic cholestasis       1.00      1.00      1.00        32
                               Common Cold       1.00      1.00      1.00        39
                                    Dengue       1.00      1.00      1.00        35
                                 Diabetes        1.00      1.00      1.00        35
              Dimorphic hemmorhoids(piles)       1.00      1.00      1.00        34
                           Drug Reaction       1.00      1.00      1.00        38
                       Fungal infection        1.00      1.00      1.00        35
                                    GERD       1.00      1.00      1.00        31
                        Gastroenteritis       1.00      1.00      1.00        36
                           Heart attack       1.00      1.00      1.00        41
                            Hepatitis B       1.00      1.00      1.00        46
                            Hepatitis C       1.00      1.00      1.00        32
                            Hepatitis D       1.00      1.00      1.00        39
                            Hepatitis E       1.00      1.00      1.00        29
                          Hypertension        1.00      1.00      1.00        33
                        Hyperthyroidism       1.00      1.00      1.00        36
                           Hypoglycemia       1.00      1.00      1.00        33
                         Hypothyroidism       1.00      1.00      1.00        30
                                Impetigo       1.00      1.00      1.00        48
                                Jaundice       1.00      1.00      1.00        36
                                 Malaria       1.00      1.00      1.00        41
                                Migraine       1.00      1.00      1.00        38
                        Osteoarthristis        1.00      1.00      1.00        38
              Paralysis (brain hemorrhage)       1.00      1.00      1.00        42
                    Peptic ulcer diseae        1.00      1.00      1.00        29
                                Pneumonia       1.00      1.00      1.00        33
                                Psoriasis       1.00      1.00      1.00        33
                              Tuberculosis      1.00      1.00      1.00        42
                                   Typhoid       1.00      1.00      1.00        33
                  Urinary tract infection       1.00      1.00      1.00        39
                           Varicose veins       1.00      1.00      1.00        35
                              hepatitis A       1.00      1.00      1.00        32

                                  accuracy                           1.00      1476
```

```
           macro avg       1.00      1.00      1.00      1476
        weighted avg       1.00      1.00      1.00      1476
```

```
# future prediction
sample = disease.sample()
sample
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chi
525	0	0	0	0	0	

1 rows × 133 columns

```
# define X_new
X_new = sample.loc[:,X.columns]
X_new
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint
525	0	0	0	0	0	0	

1 rows × 132 columns

```
# predict for X_new
model.predict(X_new)
```

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

Colab paid products - Cancel contracts here

✓ 0s completed at 5:21 PM