```
#Classification Model to Identify Multiple Disease
#Data Science internship project
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import accuracy_score

dataframe = pd.read_csv("https://raw.githubusercontent.com/ybifoundation/Dataset/main/MultipleDiseasePrediction.csv")
dataframe
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain
0	1	1	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0
2	1	0	1	0	0	0	0	0
3	1	1	0	0	0	0	0	0
4	1	1	1	0	0	0	0	0
...	...	...	...	...	...	...	...	...
4915	0	0	0	0	0	0	0	0
4916	0	1	0	0	0	0	0	0
4917	0	0	0	0	0	0	0	0
4918	0	1	0	0	0	0	1	0
4919	0	1	0	0	0	0	0	0

4920 rows × 133 columns

```
dataframe.head()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	aci
0	1	1	1	0	0	0	0	0	
1	0	1	1	0	0	0	0	0	
2	1	0	1	0	0	0	0	0	
3	1	1	0	0	0	0	0	0	
4	1	1	1	0	0	0	0	0	

5 rows × 133 columns

```
# getting some info about the data
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919
Columns: 133 entries, itching to prognosis
dtypes: int64(132), object(1)
memory usage: 5.0+ MB
```

```
# verify shape
dataframe.shape
```

```
(4920, 133)
```

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

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	s
count	4920.000000	4920.000000	4920.000000	4920.000000	4920.000000	4920.000000	4920.000000	
mean	0.137805	0.159756	0.021951	0.045122	0.021951	0.162195	0.139024	
std	0.344730	0.366417	0.146539	0.207593	0.146539	0.368667	0.346007	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	

8 rows × 132 columns

```
# check for missing value
dataframe.isnull()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
4915	False	False	False	False	False	False	False	False
4916	False	False	False	False	False	False	False	False
4917	False	False	False	False	False	False	False	False
4918	False	False	False	False	False	False	False	False
4919	False	False	False	False	False	False	False	False

4920 rows × 133 columns

```
# checking for missing values
dataframe.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
```

```
names = dataframe.columns.values
print(names)
```

```
['itching' 'skin_rash' 'nodal_skin_eruptions' 'continuous_sneezing'
 'shivering' 'chills' 'joint_pain' 'stomach_pain' 'acidity'
 'ulcers_on_tongue' 'muscle_wasting' 'vomiting' 'burning_micturition'
 'spotting_ urination' 'fatigue' 'weight_gain' 'anxiety'
 'cold_hands_and_feets' 'mood_swings' 'weight_loss' 'restlessness'
 'lethargy' 'patches_in_throat' 'irregular_sugar_level' 'cough'
 'high_fever' 'sunken_eyes' 'breathlessness' 'sweating' 'dehydration'
 'indigestion' 'headache' 'yellowish_skin' 'dark_urine' 'nausea'
 'loss_of_appetite' 'pain_behind_the_eyes' 'back_pain' 'constipation'
 'abdominal_pain' 'diarrhoea' 'mild_fever' 'yellow_urine'
 'yellowing_of_eyes' 'acute_liver_failure' 'fluid_overload'
 'swelling_of_stomach' 'swelled_lymph_nodes' 'malaise'
 'blurred_and_distorted_vision' 'phlegm' 'throat_irritation'
 'redness_of_eyes' 'sinus_pressure' 'runny_nose' 'congestion' 'chest_pain'
 'weakness_in_limbs' 'fast_heart_rate' 'pain_during_bowel_movements'
 'pain_in_anal_region' 'bloody_stool' 'irritation_in_anus' 'neck_pain'
 'dizziness' 'cramps' 'bruising' 'obesity' 'swollen_legs'
 'swollen_blood_vessels' 'puffy_face_and_eyes' 'enlarged_thyroid'
 'brittle_nails' 'swollen_extremeties' 'excessive_hunger'
 'extra_marital_contacts' 'drying_and_tingling_lips' 'slurred_speech'
```

```
 'knee_pain' 'hip_joint_pain' 'muscle_weakness' 'stiff_neck'
 'swelling_joints' 'movement_stiffness' 'spinning_movements'
 'loss_of_balance' 'unsteadiness' 'weakness_of_one_body_side'
 'loss_of_smell' 'bladder_discomfort' 'foul_smell_of urine'
 'continuous_feel_of_urine' 'passage_of_gases' 'internal_itching'
 'toxic_look_(typhos)' 'depression' 'irritability' 'muscle_pain'
 'altered_sensorium' 'red_spots_over_body' 'belly_pain'
 'abnormal_menstruation' 'dischromic _patches' 'watering_from_eyes'
 'increased_appetite' 'polyuria' 'family_history' 'mucoid_sputum'
 'rusty_sputum' 'lack_of_concentration' 'visual_disturbances'
 'receiving_blood_transfusion' 'receiving_unsterile_injections' 'coma'
 'stomach_bleeding' 'distention_of_abdomen'
 'history_of_alcohol_consumption' 'fluid_overload.1' 'blood_in_sputum'
 'prominent_veins_on_calf' 'palpitations' 'painful_walking'
 'pus_filled_pimples' 'blackheads' 'scurring' 'skin_peeling'
 'silver_like_dusting' 'small_dents_in_nails' 'inflammatory_nails'
 'blister' 'red_sore_around_nose' 'yellow_crust_ooze' 'prognosis']
```

```
dataframe.std()
```

```
<ipython-input-9-90bdd6cc73ba>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in
  dataframe.std()
itching                 0.344730
skin_rash               0.366417
nodal_skin_eruptions    0.146539
continuous_sneezing     0.207593
shivering               0.146539
                            ...
small_dents_in_nails    0.150461
inflammatory_nails      0.150461
blister                 0.150461
red_sore_around_nose    0.150461
yellow_crust_ooze       0.150461
Length: 132, dtype: float64
```

```
# check for categories
dataframe.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
```

```
dataframe.prognosis.unique()
```

```
array(['Fungal infection', 'Allergy', 'GERD', 'Chronic cholestasis',
       'Drug Reaction', 'Peptic ulcer diseae', 'AIDS', 'Diabetes ',
       'Gastroenteritis', 'Bronchial Asthma', 'Hypertension ', 'Migraine',
       'Cervical spondylosis', 'Paralysis (brain hemorrhage)', 'Jaundice',
       'Malaria', 'Chicken pox', 'Dengue', 'Typhoid', 'hepatitis A',
       'Hepatitis B', 'Hepatitis C', 'Hepatitis D', 'Hepatitis E',
       'Alcoholic hepatitis', 'Tuberculosis', 'Common Cold', 'Pneumonia',
       'Dimorphic hemmorhoids(piles)', 'Heart attack', 'Varicose veins',
       'Hypothyroidism', 'Hyperthyroidism', 'Hypoglycemia',
```

```
       'Osteoarthristis', 'Arthritis',
       '(vertigo) Paroymsal  Positional Vertigo', 'Acne',
       'Urinary tract infection', 'Psoriasis', 'Impetigo'], dtype=object)
```

```
dataframe.prognosis.nunique()
```

```
41
```

```
# correlation
#no need
```

```
# column names
dataframe.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 = dataframe['prognosis']
print(Y)
```

```
0                            Fungal infection
1                            Fungal infection
2                            Fungal infection
3                            Fungal infection
4                            Fungal infection
                        ...
4915    (vertigo) Paroymsal  Positional Vertigo
4916                                   Acne
4917                    Urinary tract infection
4918                                 Psoriasis
4919                                  Impetigo
Name: prognosis, Length: 4920, dtype: object
```

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

```
4916       0         1                   0                  0
4917       0         0                   0                  0
4918       0         1                   0                  0
4919       0         1                   0                  0

      shivering  chills  joint_pain  stomach_pain  acidity  ulcers_on_tongue  \
0             0       0           0             0        0                 0
1             0       0           0             0        0                 0
2             0       0           0             0        0                 0
3             0       0           0             0        0                 0
4             0       0           0             0        0                 0
...         ...     ...         ...           ...      ...               ...
4915          0       0           0             0        0                 0
4916          0       0           0             0        0                 0
4917          0       0           0             0        0                 0
4918          0       0           1             0        0                 0
4919          0       0           0             0        0                 0

      ...  pus_filled_pimples  blackheads  scurring  skin_peeling  \
0     ...                   0           0         0             0
1     ...                   0           0         0             0
2     ...                   0           0         0             0
3     ...                   0           0         0             0
4     ...                   0           0         0             0
...   ...                 ...         ...       ...           ...
4915  ...                   0           0         0             0
4916  ...                   1           1         1             0
4917  ...                   0           0         0             0
4918  ...                   0           0         0             1
4919  ...                   0           0         0             0

      silver_like_dusting  small_dents_in_nails  inflammatory_nails  blister  \
0                       0                     0                   0        0
1                       0                     0                   0        0
2                       0                     0                   0        0
3                       0                     0                   0        0
4                       0                     0                   0        0
...                   ...                   ...                 ...      ...
4915                    0                     0                   0        0
```

```
2                     0                 0
3                     0                 0
4                     0                 0
...                   ...               ...
4915                  0                 0
4916                  0                 0
4917                  0                 0
4918                  0                 0
4919                  1                 1

[4920 rows x 132 columns]
```

```
# 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 = 0.2, stratify=Y, random_state=2)
```

```
print(X.shape, X_train.shape, X_test.shape)
```

```
(4920, 132) (3936, 132) (984, 132)
```

```
# verify shape
len(X_train)
```

```
3936
```

```
len(X_test)
```

```
984
```

```
# select model
from sklearn.ensemble import RandomForestClassifier
classifier =RandomForestClassifier(n_estimators = 100, criterion = 'gini')
```

```
#RandomForestClassifier()
# train model
classifier.fit(X_train, Y_train)
```

```
RandomForestClassifier()
```

```
# model accuracy
# accuracy score on the training data
X_train_prediction = classifier.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
print('Accuracy score of the training data : ', training_data_accuracy)
```

```
Accuracy score of the training data :  1.0
```

```
# accuracy score on the test data
X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
print('Accuracy score of the test data : ', test_data_accuracy)
```

```
Accuracy score of the test data :  1.0
```

```
# predict with model
Y_predict=classifier.predict(X_test)
```

```
X_test
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain

```
# model evaluation
classifier.score(X_test,Y_test)
```

```
1.0
```

4512	0	0	0	0	0	0	0	1

```
# model confusion matrix
from sklearn.metrics import classification_report
```

...	...	...	...	...	...	...	...	...

```
# model classification report
print(classification_report(Y_test,Y_predict))
```

```
                                    precision    recall  f1-score   support

     (vertigo) Paroymsal  Positional Vertigo       1.00      1.00      1.00        24
                                   AIDS       1.00      1.00      1.00        24
                                   Acne       1.00      1.00      1.00        24
                    Alcoholic hepatitis       1.00      1.00      1.00        24
                                Allergy       1.00      1.00      1.00        24
                              Arthritis       1.00      1.00      1.00        24
                       Bronchial Asthma       1.00      1.00      1.00        24
                   Cervical spondylosis       1.00      1.00      1.00        24
                            Chicken pox       1.00      1.00      1.00        24
                    Chronic cholestasis       1.00      1.00      1.00        24
                            Common Cold       1.00      1.00      1.00        24
                                 Dengue       1.00      1.00      1.00        24
                               Diabetes       1.00      1.00      1.00        24
            Dimorphic hemmorhoids(piles)       1.00      1.00      1.00        24
                          Drug Reaction       1.00      1.00      1.00        24
                     Fungal infection       1.00      1.00      1.00        24
                                 GERD       1.00      1.00      1.00        24
                       Gastroenteritis       1.00      1.00      1.00        24
                          Heart attack       1.00      1.00      1.00        24
                           Hepatitis B       1.00      1.00      1.00        24
                           Hepatitis C       1.00      1.00      1.00        24
                           Hepatitis D       1.00      1.00      1.00        24
                           Hepatitis E       1.00      1.00      1.00        24
                          Hypertension       1.00      1.00      1.00        24
                       Hyperthyroidism       1.00      1.00      1.00        24
                          Hypoglycemia       1.00      1.00      1.00        24
                       Hypothyroidism       1.00      1.00      1.00        24
                              Impetigo       1.00      1.00      1.00        24
                              Jaundice       1.00      1.00      1.00        24
                               Malaria       1.00      1.00      1.00        24
                              Migraine       1.00      1.00      1.00        24
                       Osteoarthristis       1.00      1.00      1.00        24
            Paralysis (brain hemorrhage)       1.00      1.00      1.00        24
                    Peptic ulcer diseae       1.00      1.00      1.00        24
                              Pneumonia       1.00      1.00      1.00        24
                              Psoriasis       1.00      1.00      1.00        24
                           Tuberculosis       1.00      1.00      1.00        24
                                Typhoid       1.00      1.00      1.00        24
                Urinary tract infection       1.00      1.00      1.00        24
                         Varicose veins       1.00      1.00      1.00        24
                             hepatitis A       1.00      1.00      1.00        24

                                accuracy                           1.00       984
                               macro avg       1.00      1.00      1.00       984
                            weighted avg       1.00      1.00      1.00       984
```

```
# future prediction
```

```
X.shape
```

```
(4920, 132)
```

```
X.head()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	aci
0	1	1	1	0	0	0	0	0	
1	0	1	1	0	0	0	0	0	
2	1	0	1	0	0	0	0	0	
3	1	1	0	0	0	0	0	0	
4	1	1	1	0	0	0	0	0	

5 rows × 132 columns

```
X.describe()
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	ch
count	4920.000000	4920.000000	4920.000000	4920.000000	4920.000000	4920.00
mean	0.137805	0.159756	0.021951	0.045122	0.021951	0.16
std	0.344730	0.366417	0.146539	0.207593	0.146539	0.36
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.00

8 rows × 132 columns

```
# predict for X_new
X_new = X.sample()
```

```
classifier.predict(X_new)
```

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