

# kaviyadevi 20106064

## Data cleaning and preprocessing

importing libraries

```
In [1]: import numpy as np  
import pandas as pd
```

importing dataset

```
In [3]: data=pd.read_csv(r"C:\Users\user\Downloads\8_BreastCancerPrediction - 8_BreastCar  
data
```

```
Out[3]:
```

|     | id       | diagnosis | radius_mean | texture_mean | perimeter_mean | area_mean | smoothness_m |
|-----|----------|-----------|-------------|--------------|----------------|-----------|--------------|
| 0   | 842302   | M         | 17.99       | 10.38        | 122.80         | 1001.0    | 0.11         |
| 1   | 842517   | M         | 20.57       | 17.77        | 132.90         | 1326.0    | 0.08         |
| 2   | 84300903 | M         | 19.69       | 21.25        | 130.00         | 1203.0    | 0.10         |
| 3   | 84348301 | M         | 11.42       | 20.38        | 77.58          | 386.1     | 0.14         |
| 4   | 84358402 | M         | 20.29       | 14.34        | 135.10         | 1297.0    | 0.10         |
| ... | ...      | ...       | ...         | ...          | ...            | ...       | ...          |
| 564 | 926424   | M         | 21.56       | 22.39        | 142.00         | 1479.0    | 0.11         |
| 565 | 926682   | M         | 20.13       | 28.25        | 131.20         | 1261.0    | 0.09         |
| 566 | 926954   | M         | 16.60       | 28.08        | 108.30         | 858.1     | 0.08         |
| 567 | 927241   | M         | 20.60       | 29.33        | 140.10         | 1265.0    | 0.11         |
| 568 | 92751    | B         | 7.76        | 24.54        | 47.92          | 181.0     | 0.05         |

569 rows × 32 columns

To print initial rows