AIDI 1002 Machine Learning Programming — Assignment - 2

Nathan Kelber
N01429819

Part 1 Instructions

```
[NbConvertApp] Converting notebook Assignment_2_Part_1_Instructions.ipynb to markdown
[NbConvertApp] Writing 1 bytes to Assignment_2_Part_1_Instructions.md
```

Part 2 Instructions

```
[NbConvertApp] Converting notebook Assignment_2_Part_2_Instructions.ipynb to markdown
[NbConvertApp] Writing 1 bytes to Assignment_2_Part_2_Instructions.md
```

1. Exploring the Breast Cancer Wisconsin Dataset - Download the dataset from following link

Download the Breast Cancer Wisconsin (Diagnostic) Dataset from following link and answer the following questions.

```
!wget -q -O wdbc.data https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data
!head -5 wdbc.data
```

```
842302,M,17.99,10.38,122.8,1001,0.1184,0.2776,0.3001,0.1471,0.2419,0.7871,1095,9.587,153.4,0.006399,0.04904,0.05373,0.01587,0.03003,0.006193,25.38,17.33,184.6,2019,0.1622,0.6656,0.7119,0.2654,0.4601,0.1189
842517,M,20.57,17.77,132.9,1326,0.08474,0.07864,0.0869,0.07017,0.1812,0.5667,0.5435,3.398,74.08,0.005225,0.01308,0.0186,0.0134,0.0139,0.003532,24.99,23.41,158.8,1956,0.1238,0.1866,0.2416,0.186,0.275,0.08902
84300903,M,19.69,21.25,130,1203,0.1096,0.1599,0.1974,0.1279,0.2069,0.5999,0.7456,4.584,94.03,0.006150,0.04006,0.05832,0.02148,0.02267,0.003707,23.57,25.53,152.1,1709,0.1444,0.4245,0.4504,0.2430,0.3613,0.08758
84348301,M,11.42,20.38,77.58,386.1,0.1425,0.2839,0.2414,0.1052,0.2597,0.9744,0.4956,1.156,3.445,0.009110,0.07458,0.05661,0.01867,0.05963,0.009208,14.91,26.50,98.87,567.7,0.2098,0.8663,0.6869,0.2575,0.6638,0.173
84358402,M,20.29,14.34,135.1,1297,0.1003,0.1328,0.1980,0.1043,0.1809,0.5883,0.7572,7.526,95.18,0.005436,0.01308,0.0186,0.0134,0.0139,0.003532,24.99,23.41,158.8,1956,0.1238,0.1866,0.2416,0.186,0.275,0.08902
```

```
!ls -l wdbc.data
```

```
-rw-r--r-- 1 root root 125204 Jan 19 2026 wdbc.data
```

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

# Load data
df = pd.read_csv('wdbc.data', header=None)

# Shape
print('Shape:', df.shape)

# Class distribution
print('\nClass distribution:')
print(df[1].value_counts)
```

```
Shape: (569, 32)

Class distribution:
<bound_method IndexOpsMixin.value_counts of 1    M
1    M
1    M
1    M
1    M
1    M
1    B
1    B
1    B
1    B
1    B
1    B
1    B
1    B
1    B
```