Assignment 3

A certain reservoir formation is known to consist of two lithologies: dolomite (pay) and shale (non-pay). Equation A3-1 below is an application of Baye's Theorem to classify intervals in the formation as either as Dolomite or Shale; on the basis of gamma ray log measurements.

$$P(D|gamma > 60) = \frac{P(D) \times P(gamma > 60|D)}{P(D) \times P(gamma > 60|D) + P(S) \times P(gamma > 60|S)} - - - - - A3 - 1$$

Here are definitions of various terms in the equation:

- $P(D|gamma > 60)$ is the conditional probability that the interval is Dolomite given that a gamma log reading is greater than 60.
- $P(D)$ is the unconditional probability that the interval is Dolomite. It can be computed thus:
  - $P(D) = \frac{number\ of\ Dolomite\ cores}{number\ of\ all\ cores}$
- $P(gamma > 60|D)$ is the conditional probability of obtaining a gamma log reading greater than 60 in the Dolomite lithology. It is obtainable by running the line of code below:
  -
  - ```
    1 - scipy.stats.norm(loc = 0, scale = 1).cdf(0.5)
    ```
  - Note: the *loc* argument should be set equal to the mean of the data while the *scale* argument should be set equal to the standard deviation of the data.
  - The only argument to the *.cdf* method is the quantile, which in this case is 60.
  - You need to import scipy.stats before running this line.
    - Follow the tutorial (attached to the course page on Moodle) on how to install scipy before attempting to the import.
- $P(S)$ is the unconditional probability that the interval is Shale. It can be computed thus:
  - $P(D) = \frac{number\ of\ \text{Shale}\ cores}{number\ of\ all\ cores}$
- $P(gamma > 60|S)$ is the conditional probability of obtaining a gamma log reading greater than 60 in the Shale lithology. It is obtainable in the same manner as $P(gamma > 60|D)$

Once $P(D|gamma > 60)$ is computed, the interval classification can be done thus:
If $P(D|gamma > 60)$ is greater or equal to 0.5, then the interval is Dolomite; otherwise, it a Shale interval.

Write a Python script to implement Equation A3-1 and to display the resulting interval classification. Save the script as *bayes_classifier.py*, commit and push it to your GitHub repository. Submit the URL to your copy of *PET328_Class_2022* repository. Furthermore, send a pull request to the original *TTOWG/ PET328_Class_2022* repository.
    ```

You may test your script with the following data:

	Dolomite	Shale
Mean	25.8	85.2
Standard Deviation	18.6	14.9
Count	476	295