There are three datasets:

Part I

The MNIST Digit Dataset
The BBC Text Classification
The Demo Weather Dataset uploaded here

Implement classifiers using **Naive Bayes** and **Logistic Regression**. Implement the following:
Identify Hyperparamaters and select possible values for them:

- For example Learning Rates of 0.0001, 0.001, 0.01, 0.1, 1.0, 1.5
- Laplace Smoothing: 0.1, 0.5, 1.0, 10, 100

Then train your system with the above or more hyperparameters and test your models on the test sets. Plot the accuracy change based on the hyperparameter variations.

In the MNIST and BBC Datasets, try at least three feature extractions ways. At least two of them should work well. For each of your feature extraction implementations, have explanations of why you implemented them.

PS:

1. To easily read the MNIST dataset - use this library
2. BBC Dataset can be found here or Download from here
3. You should get MNIST accuracy of about 75 - 85% and on the BBC dataset about 60 - 80%

Do not use any kind of extra tools such as Numpy or Pytorch. Furthermore, do not use ChatGPT or Bard (You might use them and I may not notice it but don't ask me when you fail the exam). Use plain Python and even if you follow some tutorial or any resource, try to understand and discuss the methods with your group so that you all have a good understanding of the algorithms and their implementations. It will help you get 100% on the exam (guaranteed).

Part II - Probability Questions

1. Suppose that P(A) = 0.4, P(B) = 0.3 and P((A ∪ B)C) = 0.42. Are A and B independent?
2. Two dice are rolled. A = 'sum of two dice equals 3' B = 'sum of two dice equals 7' C = 'at least one of the dice shows a 1'

a. What is P(A|C)?
b. What is P(B|C)?
c. Are A and C independent? What about B and C?

3. Let C and D be two events with P(C) = 0.25, P(D) = 0.45, and P(C ∩ D) = 0.1. What is P(Cc ∩ D)?
4. There are 3 arrangements of the word DAD, namely DAD, ADD, and DDA. How many arrangements are there of the word PROBABILITY?
5. Let A and B be two events. Suppose the probability that neither A or B occurs is 2/3. What is the probability that one or both occur?
6. Let X denote the number of times a photocopy machine will malfunction: 0, 1, 2, or 3 times, on any given month. Let Y denote the number of times a technician is called on an emergency call. The joint p.m.f. p(x, y) is presented in the table below:

	x				
y	0	1	2	3	$p_Y(y)$
0	0.15	0.30	0.05	0	0.50
1	0.05	0.15	0.05	0.05	0.30
2	0	0.05	0.10	0.05	0.20
$p_X(x)$	0.20	0.50	0.20	0.10	1.00

- Find the probability P(Y > X).
- Find p X(x), the marginal p.m.f. of X.
- Find p Y(y), the marginal p.m.f. of Y.
- Are X and Y independent?

7. The following are data points with their labels:
   (1, 2, 3, 4), 1
   (5, 6, 7, 8), 0
   (9, 10, 11, 12), 1

   The following are the randomly set weights:

w1 = 0.1
w2 = 0.2
w3 = -0.1
w4 = 0.0

Task: make three learning updates with a learning rate of 0.1 using the data points. The updates should be based on both the Perceptron and the logistic regression. Compare the two results.