Week 1- Naive Bayes (formative assignment)

In this assignment you will perform an appropriate analysis to formulate a substantiated answer to a problem using a Naive Bayes classification model.

As part of the assignment you will create, as specified in the rubric (criteria 1 through 3):

1. *a written justification of the choice for one or more machine learning models suitable for the problem.*
2. *a detailed analysis using an existing implementation.*
3. *a written evaluation of the model's quality using an appropriate criterion.*

Work out your answers to the assignments/questions below in a Jupyter Notebook. Clearly separate natural text in Markdown cells (used for explanations, justifications, conclusions and reflections) from program code (used for analysis or visualisations).
At the same time ensure the notebook still provides a cohesive account of the analysis.

Assignment Description

1. Download and study the E-coli dataset.
2. Program a Naive Bayes classifier using sklearn to predict the location of a protein in a cell, given the characteristics described in the dataset. *(criterium 2)*
3. Evaluate the quality of the model you have created. *(criterium 3)*
4. Consider the problem of accuracy as an evaluation criterion for this dataset and work out a possible solution/improvement. *(criterium 3)*
5. Describe whether the data satisfies the naive Naive Bayes assumption. Substantiate this with statistical methods from previous years. *(criterium 1)*
6. Explain which columns should perhaps not be included when training the model. *(criterium 1)*
7. Improve the model based on your answers to questions 5 and 6. *(criterium 2)*

Assignment deliverables

Please submit the following three components:

1. **Documented implementation**
   Submit a compressed (zipped) folder containing your Python scripts or Jupyter notebooks.
   - Do not include generated output files or your virtual environment.
   - If you used a virtual environment, include a requirements.txt file listing all required libraries.

   Format: .zip
2. **Printable Version of Your Code**
   Provide a PDF version of your code (either notebooks or stand-alone scripts). This helps us offer detailed, line-by-line feedback.
   a. Ensure code readability: use landscape orientation if needed to accommodate longer lines.

   Format: .pdf
3. **Individual Contribution & AI Usage Report**
   Write a short document addressing the following points:
   a. **Individual Contribution**: Describe which parts of the assignment you completed and what responsibilities you took on.
   b. **AI Usage**: Specify whether you used AI tools and, if so, how. Be transparent about the extent and purpose of any AI support.

   Format:.pdf

Important Notes

1. Before submitting your notebook, restart the kernel and run all cells to ensure it executes cleanly from top to bottom. Remove any error messages or irrelevant outputs.
2. As with all your written assignments, you may build upon external ideas (including those from chatbots or AI tools), provided you cite your sources. Use APA or IEEE referencing style.
3. AI usage must be clearly documented. You may use AI tools as a source of feedback or inspiration, but the final work **must** be your own.