BIN2023R01 – INTRODUCTION TO DATAMINING & MACHINE LEARNING FOR BIOINFORMATICS

Lab Exercise 9- Classification with Naive Bayes

Aim: To construct a Gaussian Naive Bayes model that can classify and predict the dataset.

Procedure:

1. Gather a biological dataset comprising at least 10 features along with a corresponding label from Kaggle.
2. Import essential packages for building and evaluating machine learning models using scikit learn.
3. Examine the distribution of data to understand its characteristics and perform necessary preprocessing steps
4. Develop a Gaussian Naive Bayes classifier model using the preprocessed dataset.
5. Assess the model's performance using appropriate evaluation metrics.
6. Utilize the trained model to make predictions on custom user-defined data.
7. Compare the performance metrics of the Gaussian Naive Bayes model with those of alternative classification models to determine the model with the highest accuracy.

Questions:

1. What is Naive Bayes classification, and how does it work? What do naïve and bayes define?
2. Explain the assumption of the Naive Bayes classifier.
3. What are the different types of Naive Bayes classifiers, and when should each be used?
4. How does the Naive Bayes algorithm handle continuous and categorical features?
5. Discuss the strengths and weaknesses of the Naive Bayes algorithm.
6. Can Naive Bayes handle missing values in the dataset?
7. When should Naive Bayes be preferred over other classification algorithms?

Soft copy deadline: April 1st, 11:59PM
Hard copy deadline: April 2nd, 3:15PM