DSC Assignment - classification

Task 1 – Bayes' theorem

A. 60% of the kids play football, and 36% of the kids play ice hockey. 40% of the kids who play football also play ice hockey. What percent of those that play ice hockey also play football?

B. 40% of the kids like music, and 24% of the kids like to dance. Given that 30% of those that like music also likes to dance, what percent of those that like to dance also likes music?

C. In a factory, machine X produces 60% of the daily output and machine Y produces 40% of the daily output.
2% of machine X's output is defective, and 1.5% of machine Y's output is defective.
One day, an item is inspected at random, and found to be defective. What is the probability that it was produced by machine X?

Task 2

Make a KNN classifier on the IRIS dataset using Python. Make sure to split the dataset into training and testing sets.

Task 3

Download KNIME. Make a KNN classifier on the IRIS dataset.

Task 4

Find a dataset that interests you on the UCI Machine Learning Repository. Make a KNN classifier on the dataset, either in Python or in KNIME. Visualize the classified data.

Task 5

Implement a Gaussian Naïve Bayes classifier on the same data set (in either Python or KNIME).

Task 6

Decide which of the two classifiers you would use on new data. What is behind your decision?