**High School Competition- Artificial Intelligence/ Machine Learning Track**

**Introduction:**

Machine learning is an application of artificial intelligence (AI) that provides the machines the ability to automatically learn from example without being explicitly programmed.

Classification is a supervised machine learning technique that is used to train machines using labeled examples in order to be able to predict future events. There are several algorithms used for classification such as decision trees, Naïve Bayes, support vector machines, and k-nearest neighbor. These algorithms can be trained on sufficient labeled samples and then they are able to make predictions about new data. When making predictions, these learning algorithms compare the predicted output (i.e., labels) with the true data labels and find prediction errors in order to modify the models and enhance the models accordingly.

Examples.py provides several explained examples using different classification models. The following are sample questions intended to help you prepare for the competition.

**Sample Questions:**

**Q1.** Write a code to create a DecisionTreeClassifier (refer to Example 1 in Examples.py), train it on *subset2*, and test it on *subset1*. Print the prediction accuracy.

**Q2.** Write a code to create a new array called *newTestArray* having the following two features [-4, -2], [0, 1]. Create a GaussianNB classifier (refer to Example 2 in Examples.py), train it on *subset1*, and test it on *newTestArray*. Print the predicted labels for the *newTestArray*.

**Q3.** Consider the data you have in the arrays *features* and *labels* (in Section 1 in Examples.py). Write a code to split each of these arrays into two random subsets: train subset that contains two-thirds (67%) of the total number of samples, and test subset that contains one-third (33%) of the total number of samples (refer to Example 4 in Examples.py). In your code, create an SVC classifier (refer to Example 3 in Examples.py), train it on the train subset, and test it on the test subset. Print the prediction accuracy.

**Note:** Solutions to these questions are available in:

SampleQuestionsSolutions.py

**Instructions:**

* To run the python (.py) files provided, you need python and sklearn toolkit installed on your PC. Instead, you can copy and paste the contents of these files in online python environments such as:

<https://repl.it/languages/python3>

<https://paiza.io/en/projects/new?language=python3>

* Ignore the run-time warnings that you may encounter because of version differences.
* The files Examples.py and SampleQuestionsSolutions.py will be available during the competition for your reference.