Date:18/09/2022

Assignment No. 05

Title: Implement the different Classifiers

NAME: Saurabh Nagre

PRN: 2019BTECS00080

**Objective/Aim:** To perform following tasks:

Design and implement the following classifiers:

a) Regression classifier.

*b) Naïve Bayesian Classifier.*

*c) k-NN classifier (Take k = 1,3,5,7)*

*d) Three layer Artificial Neural Network (ANN) classifier (use back*

*propagation). Plot error graph (iteration vs error).*

**Introduction:**

A classifier in machine learning is an algorithm that automatically orders or categorizes data into one or more of a set of “classes.” One of the most common examples is an email classifier that scans emails to filter them by class label: Spam or Not Spam.

**Theory/Algorithm:**

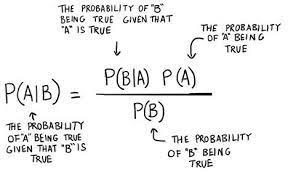
**Logistic Regression:**

This type of statistical model (also known as logit model) is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn’t vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure. This is also commonly known as the log odds, or the natural logarithm of odds, and this logistic function is represented by the following formulas:

Logit(pi) = 1/(1+ exp(-pi))

**Naive Bayes Classifier:**

Naive Bayes is a family of probabilistic algorithms that calculate the possibility that any given data point may fall into one or more of a group of categories (or not). In text analysis, Naive Bayes is used to categorize customer comments, news articles, emails, etc., into subjects, topics, or “tags” to organize them according to predetermined criteria.



Naive Bayes algorithms calculate the probability of each tag for a given text, then output for the highest probability:

Meaning, the probability of A, if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true.

Moving from tag to tag, this calculates the probability that a data point belongs within a certain category or not: Yes/No.

**K-Nearest Neighbors**

K-nearest neighbors (k-NN) is a pattern recognition algorithm that stores and learns from training data points by calculating how they correspond to other data in n-dimensional space. K-NN aims to find the k closest related data points in future, unseen data.

In text analysis, k-NN would place a given word or phrase within a predetermined category by calculating its nearest neighbor: k is decided by a plurality vote of its neighbors. If k = 1, it would be tagged into the class nearest 1.

**ANN:**

Artificial neural networks are designed to work much like the human brain does. They connect problem-solving processes in a chain of events, so that once one algorithm or process has solved a problem, the next algorithm (or link in the chain) is activated.

Artificial neural networks or “deep learning” models require vast amounts of training data because their processes are highly advanced, but once they have been properly trained, they can perform beyond other, individual, algorithms.

There are a variety of artificial neural networks, including convolutional, recurrent, feed-forward, etc., and the machine learning architecture best suited to your needs depends on the problem you’re aiming to solve.

**Procedure:**

**Import Data file**

**Choose**

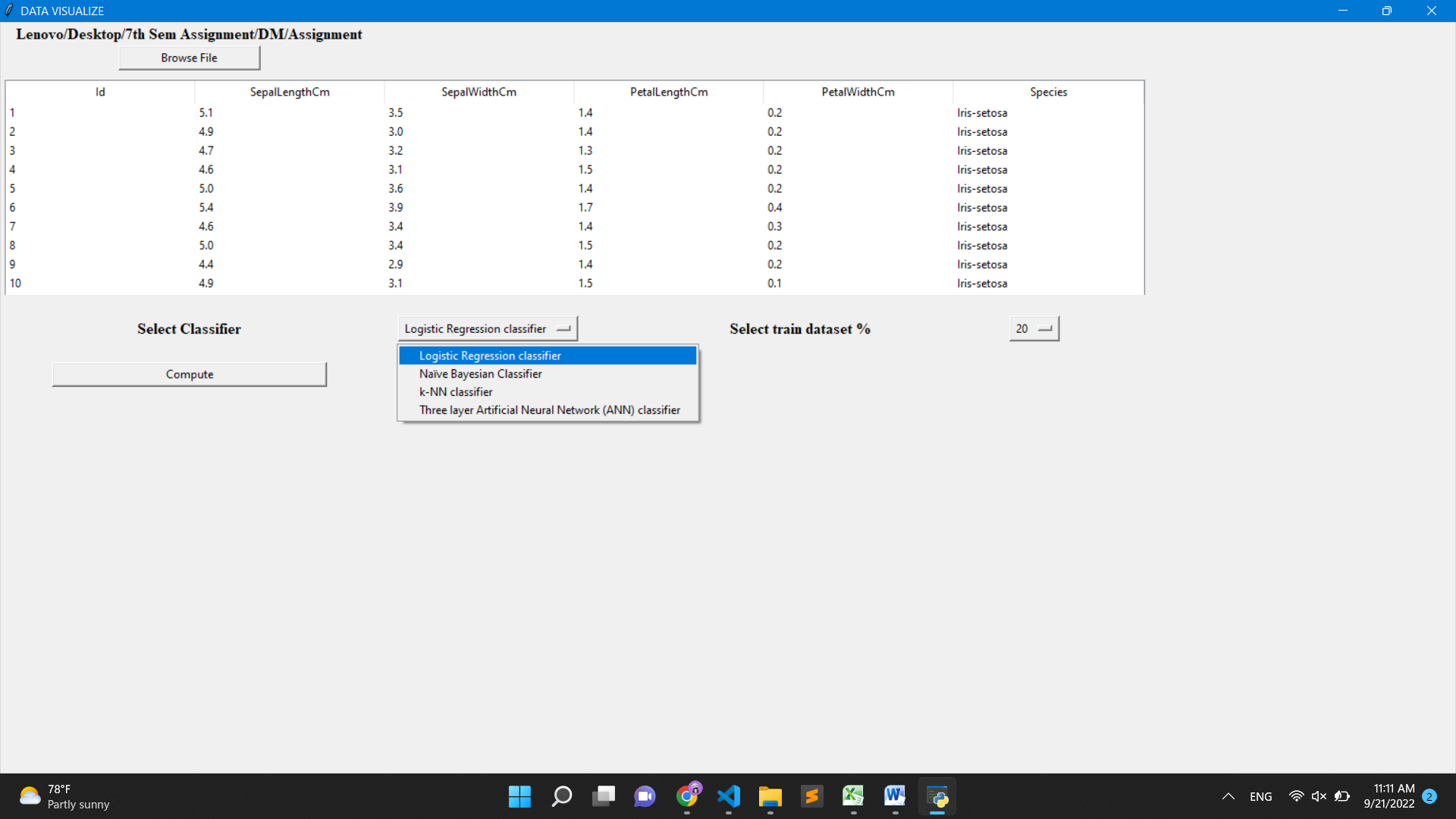
**Classifier & Train%**

**Implement Classifier**

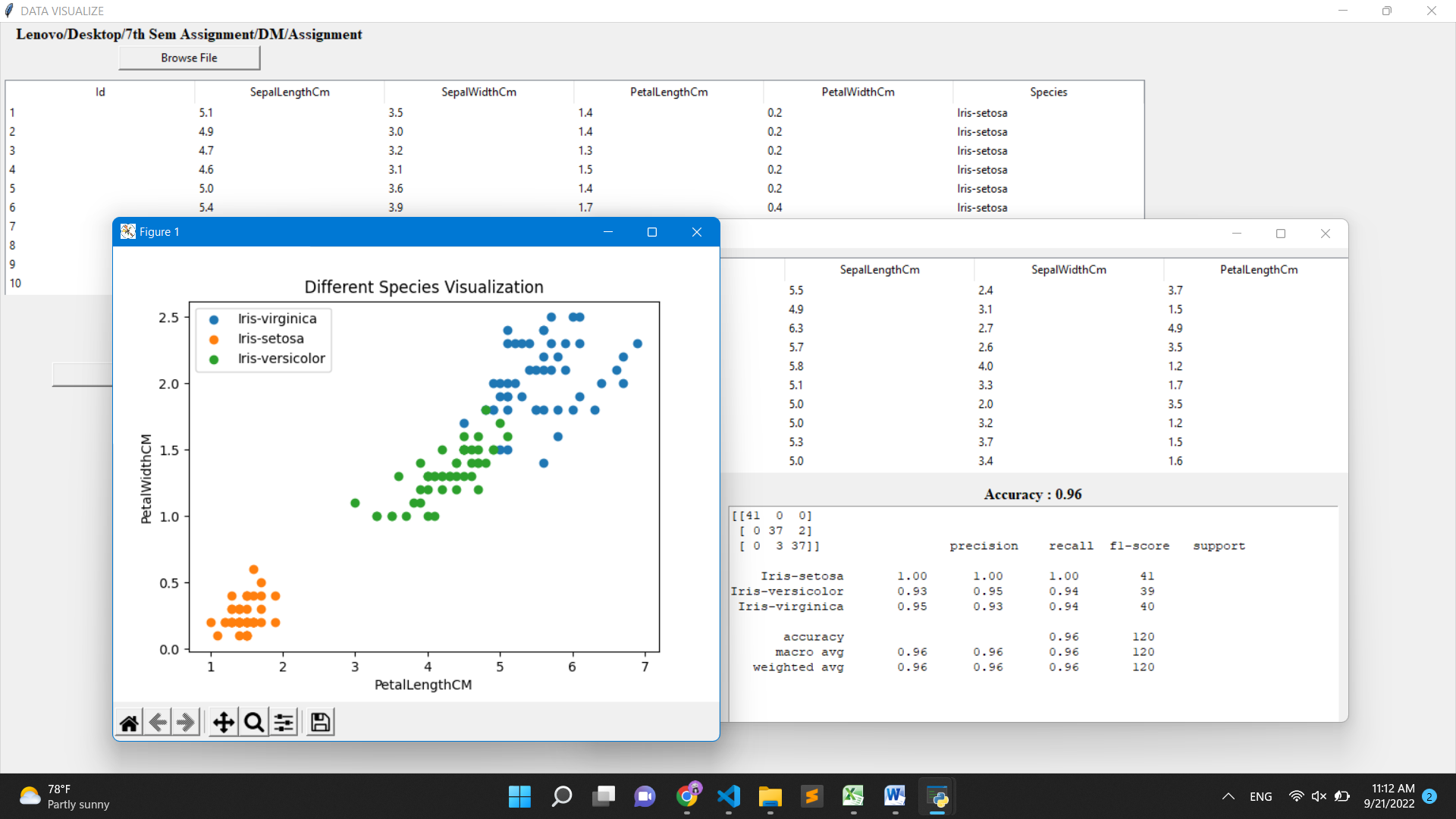
**View Result**

**Actual Experimentation/ simulation/ result/ Observation:**

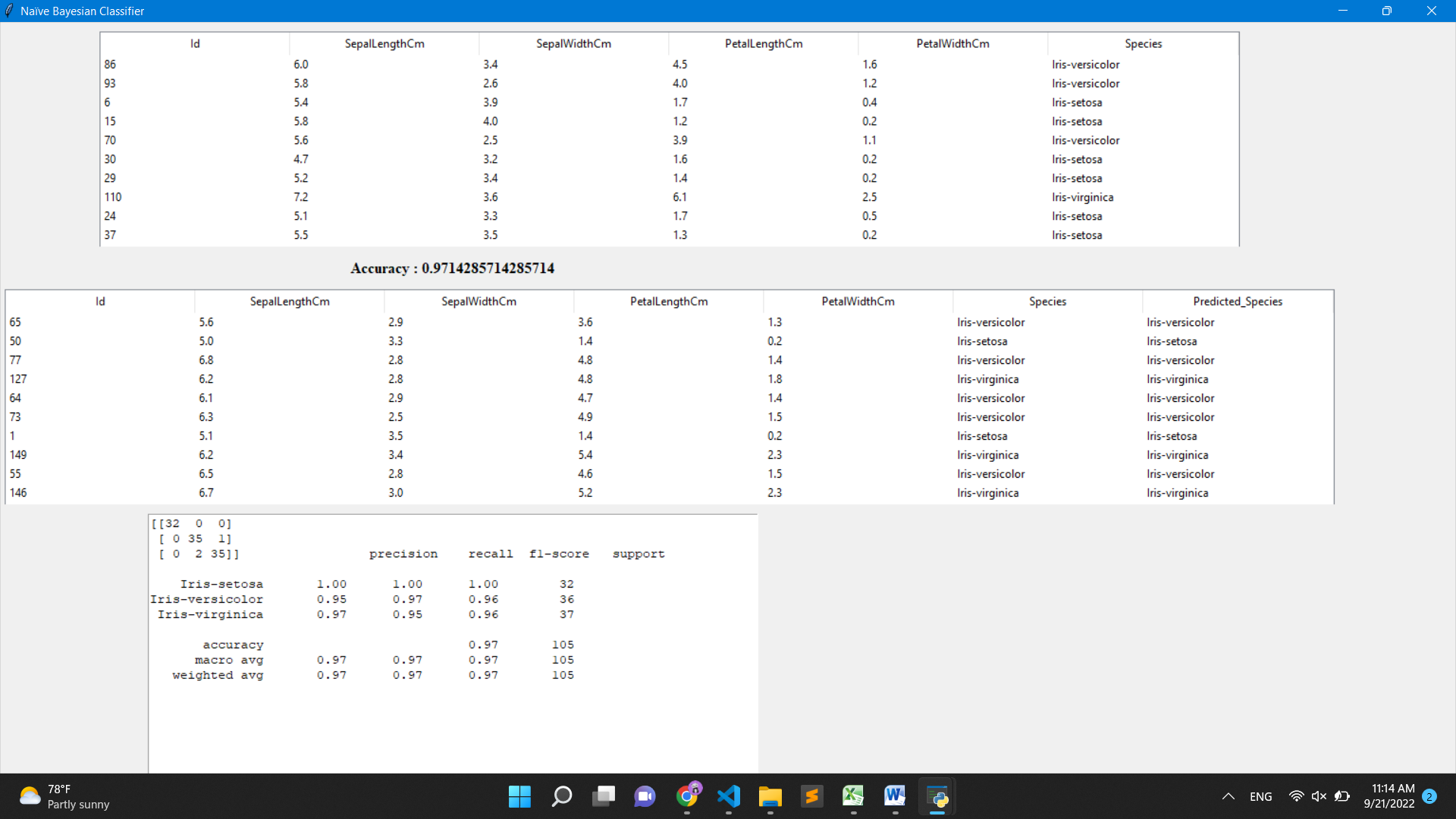
The GUI:



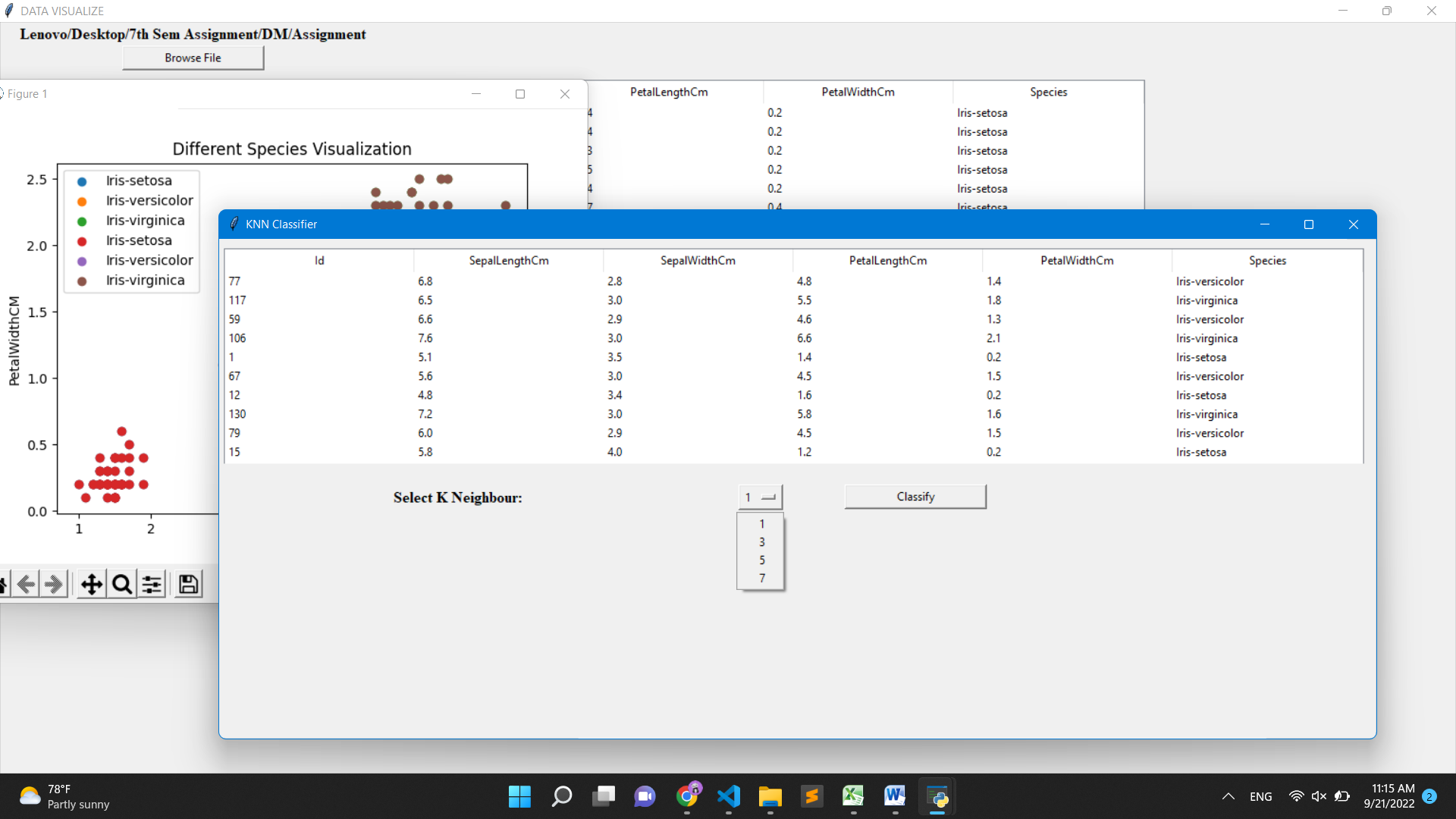
Logistic Regression:

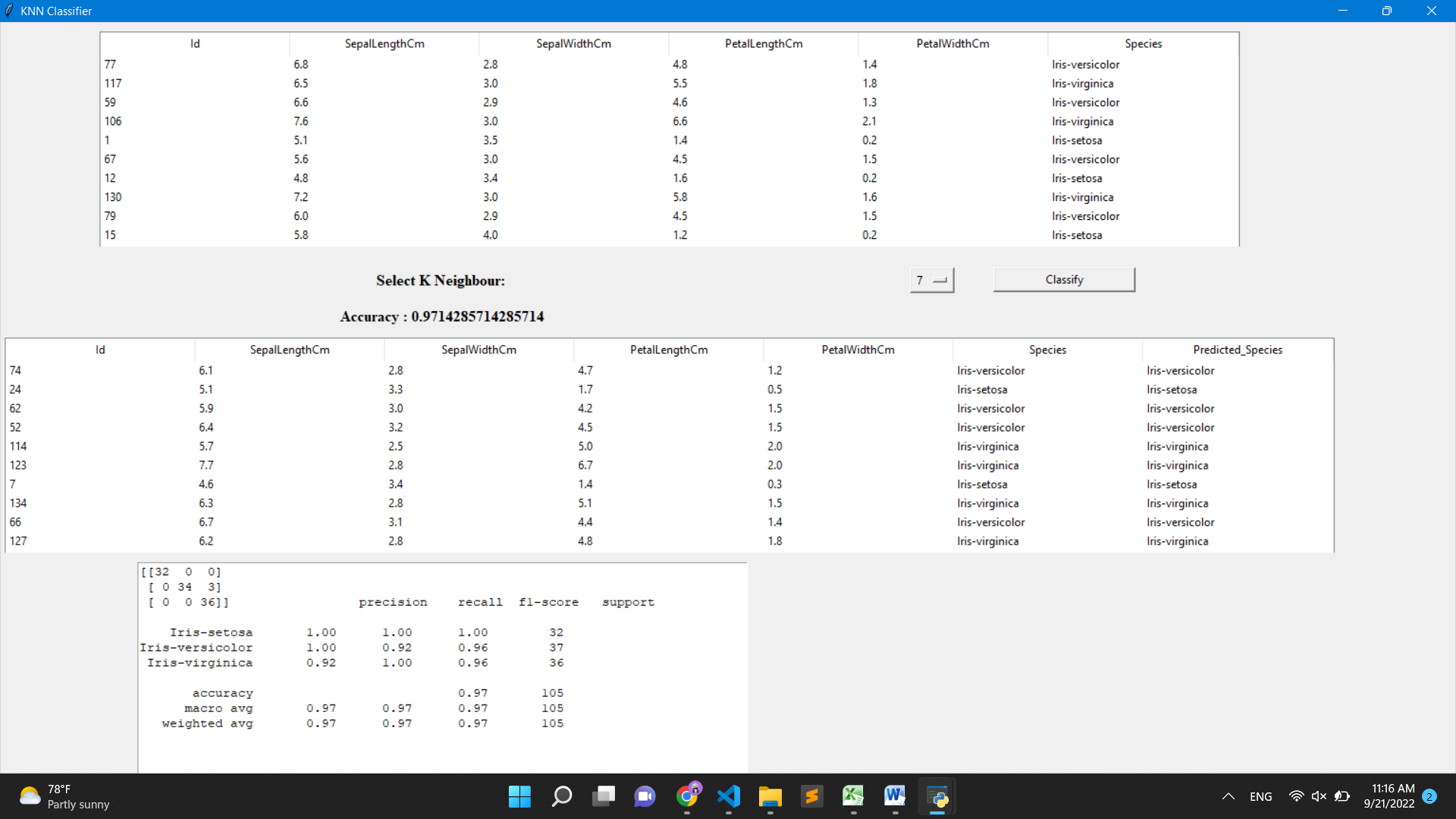


Naïve Bayesian Classifier:

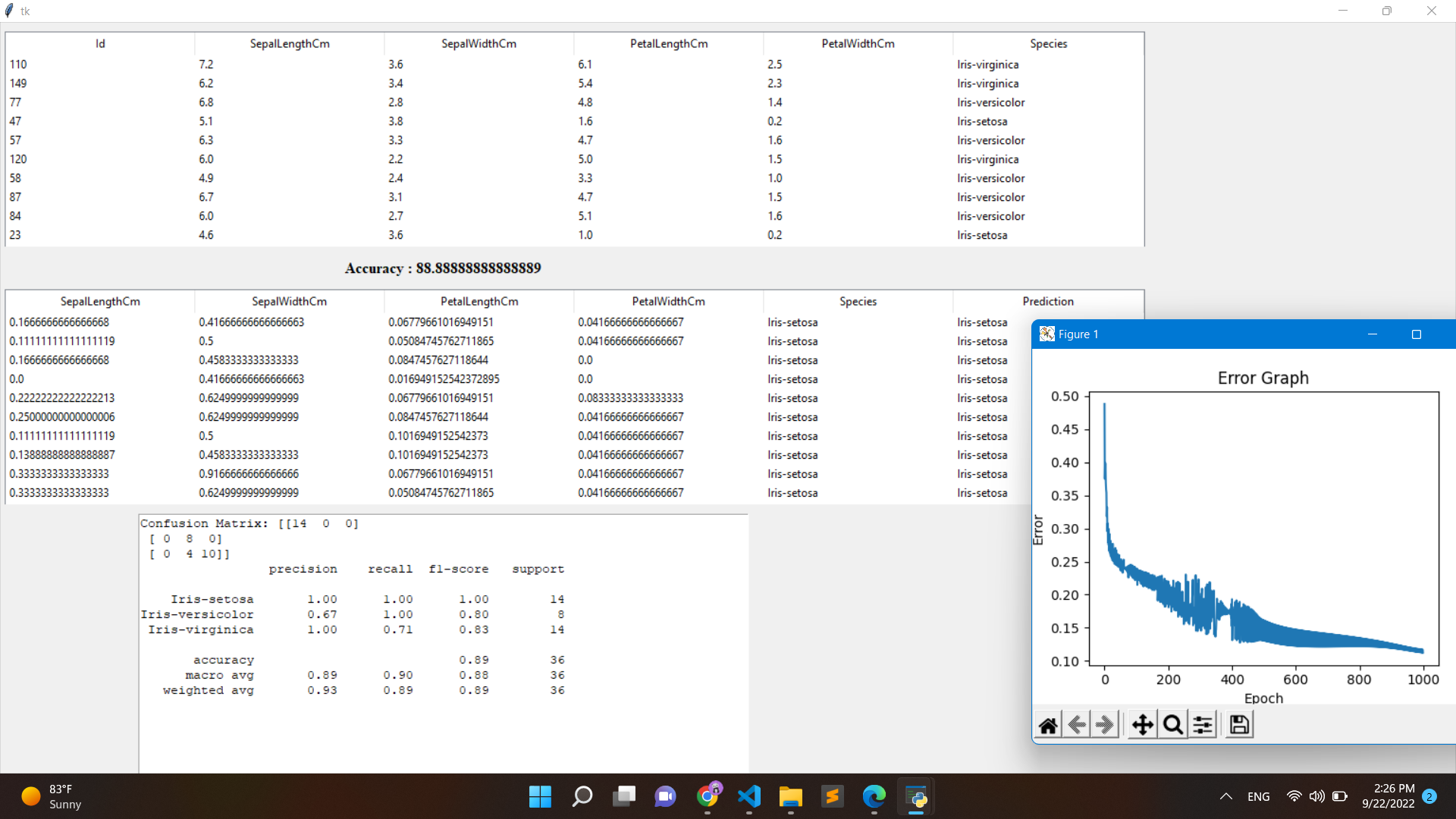


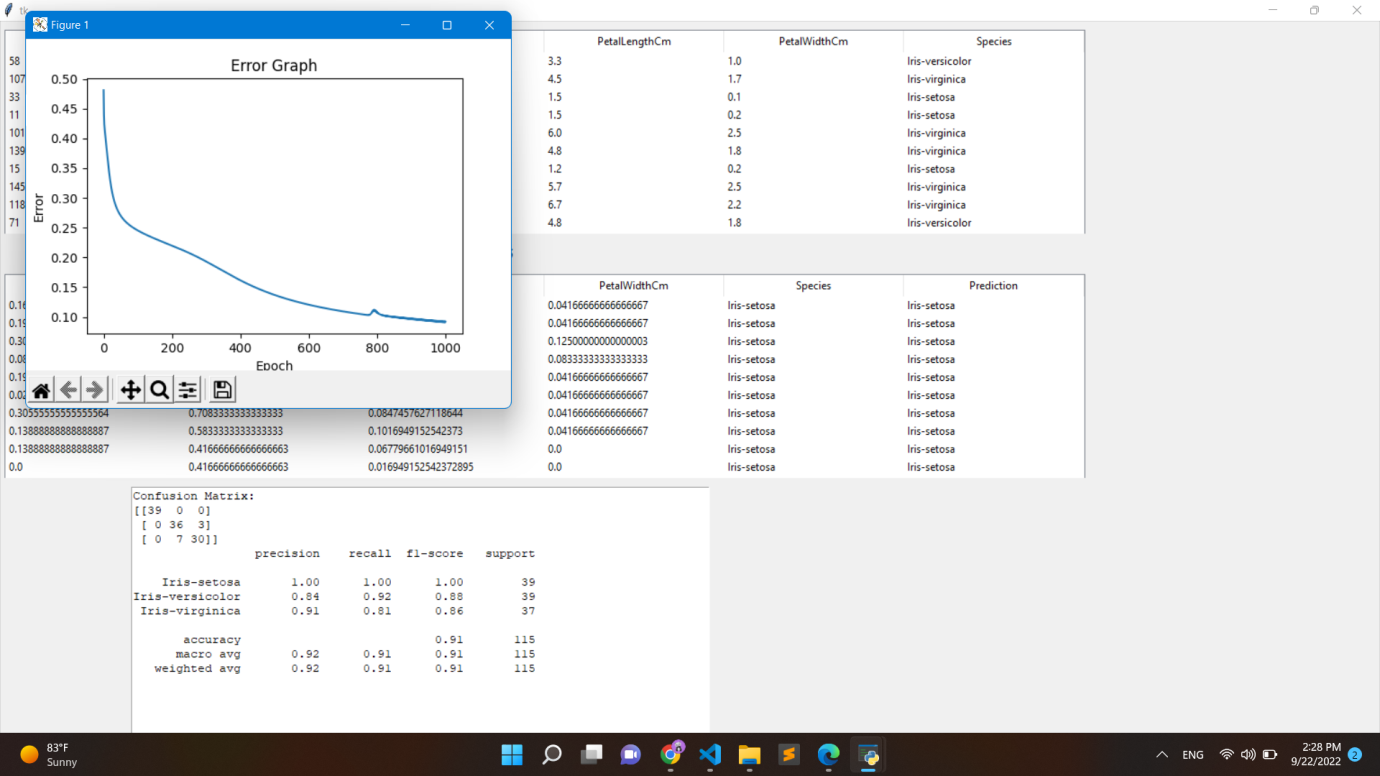
KNN classifier:





ANN :





**Conclusion:**

We can use various machine learning classifier to classify objects. We can improve our accuracy of model providing proper preprocessing and providing more training data.

**References:**

1. [https://monkeylearn.com/blog/what-is-a-classifier/#:~:text=A%20classifier%20in%20machine%20learning%20is%20an%20algorithm%20that%20automatically,label%3A%20Spam%20or%20Not%20Spam.](https://monkeylearn.com/blog/what-is-a-classifier/%23:~:text=A%20classifier%20in%20machine%20learning%20is%20an%20algorithm%20that%20automatically,label%3A%20Spam%20or%20Not%20Spam.)
2. [https://www.kaggle.com/code/burhanykiyakoglu/k-nn-logistic-regression-k-fold-cv-from-scratch#k-NN,-Logistic-Regression-and-k-Fold-Cross-Validation-from-Scratch](https://www.kaggle.com/code/burhanykiyakoglu/k-nn-logistic-regression-k-fold-cv-from-scratch%23k-NN,-Logistic-Regression-and-k-Fold-Cross-Validation-from-Scratch)
3. <https://www.kaggle.com/code/antmarakis/another-neural-network-from-scratch>