Pattern Recognition Course Project

**Mina Sherif Henry**

# Environment

Language: Python 3.6

IDE: Pycharm

Machine Learning Library: Scikit-learn

Github Link: <https://github.com/mina37/PatternRecognition>

# Dataset

Titanic Dataset

<https://www.kaggle.com/c/titanic/data>

# Classification

Based on the data 4 features were used to teach four different algorithms and make them after that predict if a different Titanic passenger has survived or not.

The features being:

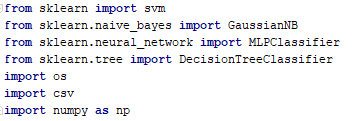
Sex(Gender), Age, SibSp(Sibling or Spouse) and Parch(Parents or Children)

Also to give everything a numerical value the Gender what given 0 for male and 1 for female. Any missing age data for any passenger was automatically replaced by 25. However at this point if all the available ages were averaged will give better results probably.

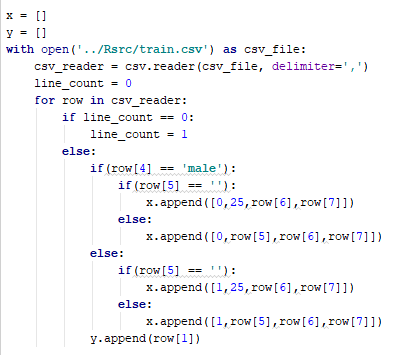
Classification algorithms used: **SVM, Naïve Bayes, Multi-layer Neural Network, Decision Tree.**

# Code and Results

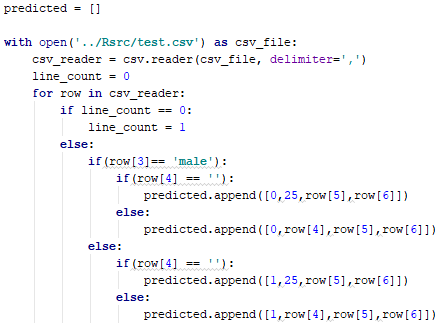
The code uses 2 main libraries: Scikit-learn and Numpy



We read the features of the training data in a 2- dimensional array “x” and the class that these features belong to – Certain passenger being survived or not – in an array “y”

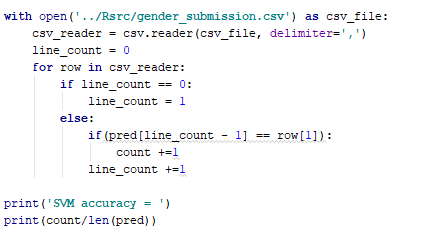


An array “Predicted” was filled with features of the test data



## SVM

SVM was fitted with default configuration. And Tested on the Test Data

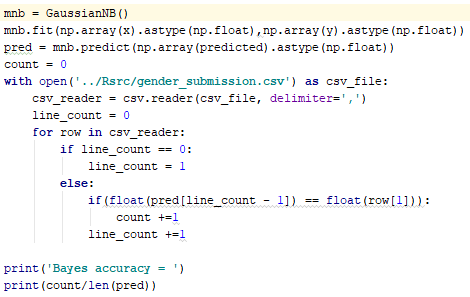


And acquired results of 91%



## Bayes

Then Gaussian Naïve Bayes was fitted with the training data. Also with default configurations and was tested with the test data



Got an accuracy of 98.8%



## Neural Network

Multi-Layer neural network was fitted with the data. However for the sake of testing it, we decided to try different number of perceptrons in the hidden layers. In the model used the number of perceptrons affected the number of layers.

Hidden Layer Sizes (2)



Hidden Layer Sizes (6)



Hidden Layer Sizes (8)



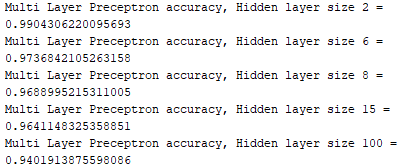
Hidden Layer Sizes (15)



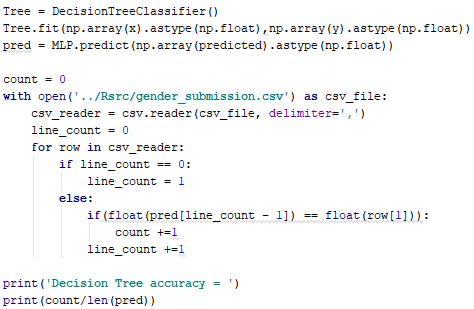
Hidden Layer Sizes (100)



Results Acquired



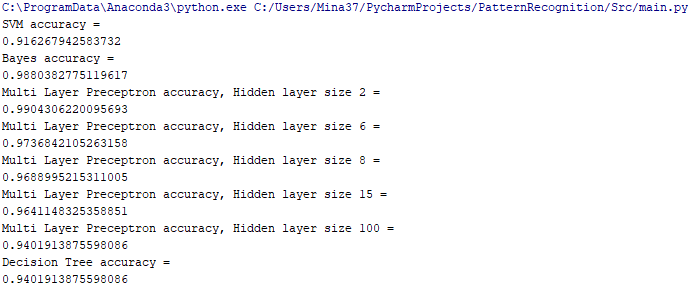
## Decision Tree



With results



# Overall Results and Conclusion



Conclusion:

The multi-layer neural network with the first configuration of 2 perceptrons in the hidden layers gave the best results of 99% accuracy