**Applied Artificial Intelligence Project**

Implemented Logistic Regression algorithm on Titanic dataset for predicting the survivors based on various features.

**Dataset used:** Titanic dataset for passenger survival prediction

Using the training dataset(train.csv) containing various features like passenger gender, age etc. as inputs and survival as the output feature, to predict the survival of passengers listed in the testing dataset(test\_input.csv). The expected output(test\_expected\_output.csv) can be compared with the algorithm's prediction.

**Algorithm used:** Logistic Regression

This is a classification algorithm which is used to analyse a dataset where there are one or more independent variables, the ones that influence the dependent variables and that help to predict an outcome. The dependent variable is binary, there are only two classes for example 0 and 1. The goal of this algorithm is to find the best fitting model by choosing the parameters that describe the relationship between the dependent and the independent variables.

**Libraries:**

Scikit learn for using the Logistic Regression algorithm

pandas for reading the training and testing data files to analyse the dataset

**Tools:** Built using Jupiter Notebook which helps to define code as well as mention comments in the file itself.

**Implementation:** Step-by-step explained below

1. Imported libraries to be used for coding and analysing the data
2. As there are many blank values in the dataset files, firstly a data clean-up on the column values obtained from the files is done before proceeding with the analysis of the data. Also, to make the values algorithm-friendly, replaced text values with numerical representations so that the algorithm can optimize the model better.
3. Selected features(independent variables) as input and set the target value(dependent variable) field to be fed into the algorithm as input and output features from the training dataset(train.csv).
4. Using linear\_model, LogisticRegression method is called and the features are fed into the instance created. To compute the accuracy of the classifier instance, the score is computed. The number of input parameters are also altered to check the variation in the score.
5. Now, using the testing dataset(test\_input.csv) the classifier is fed the input features so that it can predict the output based on the learning. The prediction values can be verified using the predict method on the classifier instance.
6. The expected output is compared against the result and the score is computed.