

# Project Proposal

Ahmed Amine Daou - Alejandra Jimenez - Gabriel Lagubeau

November 18, 2018

The aim of this work is to compare the performances of three different machine learning algorithms. The performance tested in this study is classification error,  $R^2$ , F test, over-fitting /under-fitting problem and curse dimensionality problem. To do that, we have chosen to test Naive-Bayes, Decision Trees and MLP in two different classification problems with large databases. First, we will use "Census Income" dataset to do a binary classification task (predicting if income exceeds 50K/yr or not), at second, we will use the CIFAR-10 dataset to do a multi-classification task (predicting each image class).

## Classifiers

- **Multilayer Perceptron:**

A Multilayer Perceptron (MLP) is a neural network that generates a set of outputs from a set of inputs. it consists of layers connected in one-way graph where Each node of hidden & output layers has a non linear activation function. In MLP we use backpropagation as a supervised learning technique.

- **Naive Bayes:**

A naive Bayes classifier is an algorithm that uses Bayes' probability theorem to classify data, what makes it different is that the probability of an event can be adjusted as new data is introduced.

- **Random Forest**

Random forests algorithm operates by constructing a number of decision trees during the training on different data subsets, then outputs class mode in classification case, or mean prediction in regression case of the individual trees. This algorithm combines the concepts of random subspaces and bagging.

## Datasets

- [The CIFAR-10 dataset](#) : A 10-class (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck ) database of 60000 32x32 color images ,50000 for training divided in five 10000 batches and 10000 for testing selected randomly from each class.
- [Adult Data Set](#) : Used to predict whether a person makes over 50K a year, relying on 48842 instances (32561 for training and 16281 for testing) having 14 attributes .

## Sources

- Multilayer Perceptron [Wikipedia](#) [Techopedia](#)
- Naive Bayes [Wikipedia](#) [Techopedia](#)
- Random Forest [Wikipedia](#) [Techopedia](#)