

Jeu de données pour le projet de cours MGL7811

Objectif : Chaque groupe devrait explorer les jeux de données et en sélectionner un pour leur projet de cours.

Classification

Jeu de données	Description (anglais)	No d'Instances	Type du problème
Spambase	Classify emails as spam, and non-spam	4,601	Classification binaire
Heart disease	Predict the risks of heart disease of patients	303	Classification binaire
Census data	Predict whether income exceeds \$50K/y based on census data.	48,842	Classification binaire
Mushroom	Classify mushrooms into poisonous or edible based on their physical characteristics	8124	Classification binaire
Obesity Levels	Estimation of obesity levels in individuals based on their habits and physical condition	2,111	Classification multi-classes
Predict Student Dropout	Predict student dropout from academic institutions based on student profiles	4,424	Classification binaire
Handwritten digits	Dataset of vectorized images of handwritten digits	5620	Classification multi-classes (optique)

Classification d'Images

Dataset	Description	# of Instances	Type of problem
ImageNet	Classify of objects in images	150,000	Classification multi-classes Reconnaissance d'objet
CIFAR-10	Image classification (32x32)	60,000	Classification multi-classes