

FOREST COVER TYPE PREDICTION

Kevin Zagalo

kevin.zagalo@etu.upmc.fr

Ismail Benkirane

ismail.benkirane@etu.upmc.fr

Projet pour le cours *Apprentissage Statistique* du LIP6, Sorbonne Université

Janvier 2019

Ce projet a pour but de proposer et tester des modèles pour l'étude de la base de données *Covertypes*¹, de 581 012 instances, avec 54 attributs et 7 classes à prédire, sans données manquantes.

Les attributs sont les suivants :

Nom	Unité	Description
Elevation	mètres	Altitude
Aspect	degrés	Orientation
Slope	degrés	Pente
Horizontal_Distance_To_Hydrology	mètres	Distance horizontale au point d'eau le plus proche
Vertical_Distance_To_Hydrology	mètres	Distance verticale au point d'eau le plus proche
Horizontal_Distance_To_Roadways	mètres	Distance horizontale à la route la plus proche
Hillshade_9am	entier entre 0 et 255	Ombrage à 9h au solstice d'été
Hillshade_Noon	entier entre 0 et 255	Ombrage à 12h au solstice d'été
Hillshade_3pm	entier entre 0 et 255	Ombrage à 15h au solstice d'été
Horizontal_Distance_To_Fire_Points	mètres	Distance horizontale au départ de feu le plus proche
Wilderness_Area	4 colonnes binaires	Wilderness area designation
Soil_Type	40 colonnes binaires	Type de sol
Cover_Type	entier entre 1 et 7	Classe

Il s'agit donc d'un problème de classification multi-classe avec 7 classes.

TABLE DES MATIÈRES

1	Chargement des données	2
2	Analyse préliminaire et pré-traitement des données	2
3	Test de différents modèles	2

1. <https://archive.ics.uci.edu/ml/datasets/Covertypes>

1 CHARGEMENT DES DONNÉES

2 ANALYSE PRÉLIMINAIRE ET PRÉ-TRAITEMENT DES DONNÉES

3 TEST DE DIFFÉRENTS MODÈLES