# Machine Learning - SS 2021
# Exercise 1: Classification

Federico Ambrogi, e1449911@student.tuwien.ac.at
Adam Höfler, e11847620@student.tuwien.ac.at
Matteo Panzeri 12039996@student.tuwien.ac.at
TU Wien

April 21, 2021

## Contents



## Introduction

In this document we describe the results of the implementation of three algorithms to solve classification problems on four different dataset. Blah

## 1 Data Sets Description

Here we briefly introduce our four datasets.

**Drug consumption data set** This is the data set taken from exercise 0, so here we only provide basic details. The primary task of our implementation is to classify the likeliness of the frequence of usage of a certain type of drug (from never used to current usage), given personal aspects such as gender, ethnicity, education background etc. and psychological traits.

**Asteroids data set** The asteroids data set is designed to perform a binary classification task. It consists of a total of XXX features, mostly astrophysical parameters of asteroids such as the distance from the Earth, the size of the major axis of the orbits, their mass etc. as well metadata regarding their name. The target feature for classification is a string variable that maps to boolean values "True" in the case the asteroid represents a concrete hazard for the Earth i.e. with high risk of impact, and "False" otherwise.

Blah

Blah

# 2 Data Exploration and Pre-processing

In this section we describe the necessary preliminary steps to import and prepare the data to make them suitable for the learning algorithms. The steps are handled by the script *data_preparation.py* which includes dedicated function to pre-process each data set.

# 3 Holdout and Cross Validation

# 4 Performance Tests

For each classifier, we produce a confusion matrix where each entry $i, j$ corresponds to the number of observations in group $i$, but predicted to be in group $j$. We chose to normalize the entries according to the sum of each row. In case of binary classification, the matrix reduces to the number of true negatives, false positives, false negatives and true positives.

Examples can be see in Fig. 1.

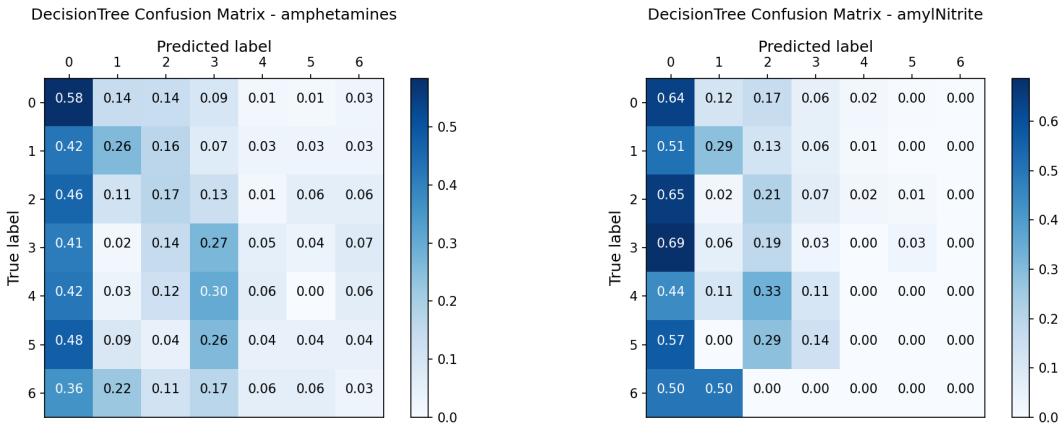


Figure 1: Example confusion matrices for XXX