



MSC. DATA SCIENCE & ARTIFICIAL INTELLIGENCE

INTRODUCTION TO MACHINE LEARNING

Dr. Michel Riveill & Dr. Diane LINGRAND

Final project: Petfinder

Author: Joris LIMONIER

joris.limonier@hotmail.fr

Due: January 15, 2022

Contents

1	Exploratory Data Analysis	1
2	Problem description	1
3	Solution	1
4	Evaluation & critical view	2

List of Figures

List of Tables

1	Accuracies of first prospect	1
2	Accuracies of first prospect	1

1 Exploratory Data Analysis

2 Problem description

Type	Column name
CATEGORICAL	Type
	Gender
	Breed
	Color1
	Color2
	Color3
	MaturitySize
	FurLength
	Vaccinated
	Dewormed
	Sterilized
	Health
NUMERICAL	Age
	Fee
TEXT	Description
IMAGE	Images

Table 1: Accuracies of first prospect

3 Solution

XGBoost is too slow.

Classifier	Accuracy
GradientBoostingClassifier	0.629
RandomForestClassifier	0.623
AdaBoostClassifier	0.612
MLPClassifier	0.602
BernoulliNB	0.600
GaussianNB	0.567
DecisionTreeClassifier	0.559
SVC	0.529
KNeighborsClassifier	0.520
GaussianProcessClassifier	0.509
SGDClassifier	0.509

Table 2: Accuracies of first prospect

4 Evaluation & critical view