Mini project : Challenge Areal Team Sputnik

Members :

VERRIEZ CHASTANG Mathilde

DORDAIN Jean-Camille

HERBERA Mathis

ZENAIDI Lydia

DAHASSE Rayane

HUART Kevin

URL of challenge :

https://codalab.lri.fr/competitions/399

Repo GitHub of project :

https://github.com/grpSputnik/Sputnik

Introduction

Context and description of the problem

* 13 category

* 9100 satellite images

* Size : 128*128

* 700 images by class

* 2 parts : preprocess and row data

Motivations

Fast description of classes

Preprocessing

* Load and transform raw data into understandable data

* 9100 images raw

* Images of validation, training and tests

* Algorithm of feature selection

* RFE

* PCA

Feature selection (before)

0	0	1
0	1	1
0	0	0
0	1	1
1	1	0
0	1	1

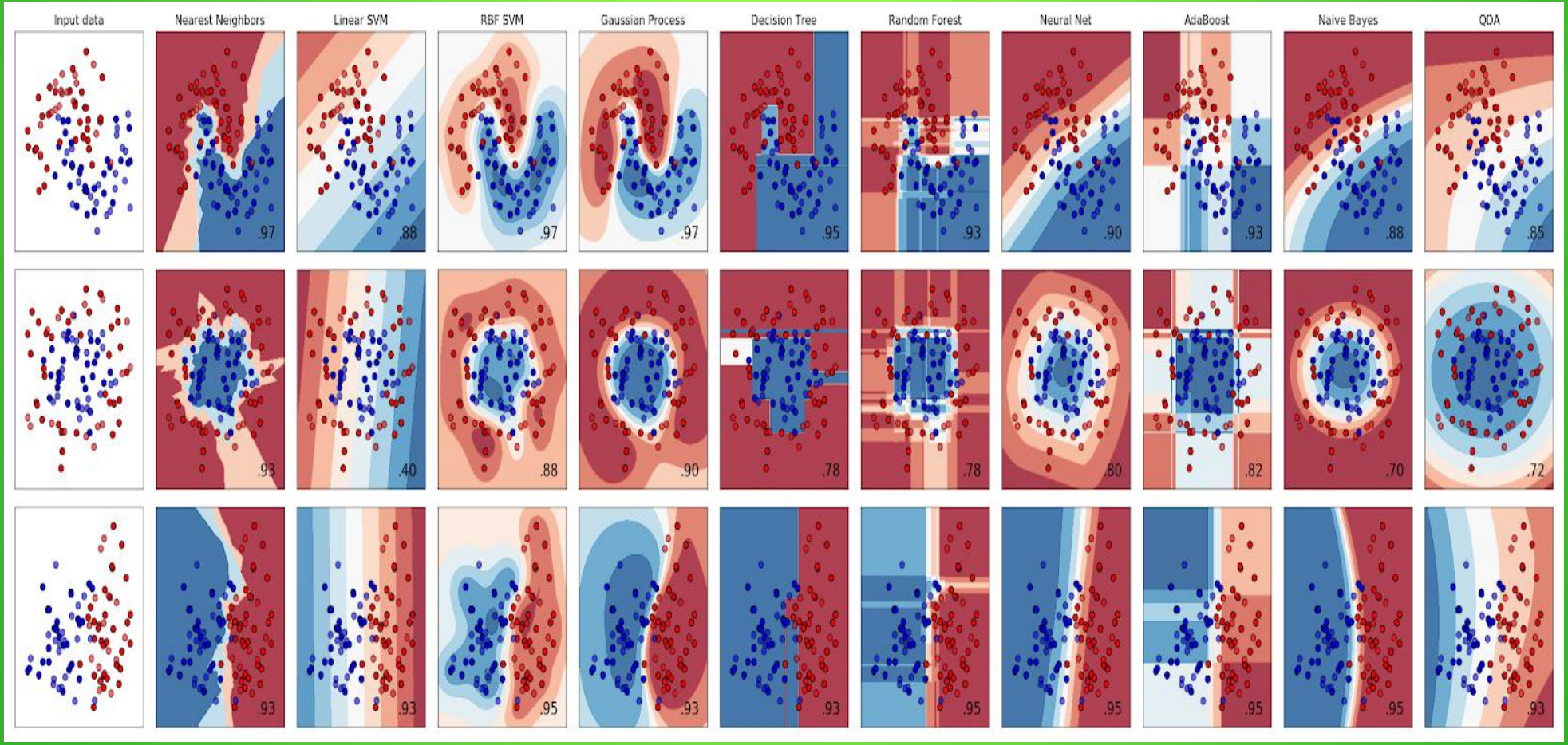
Feature selection (after)

0	1
1	1
0	0
1	1
1	0
1	1

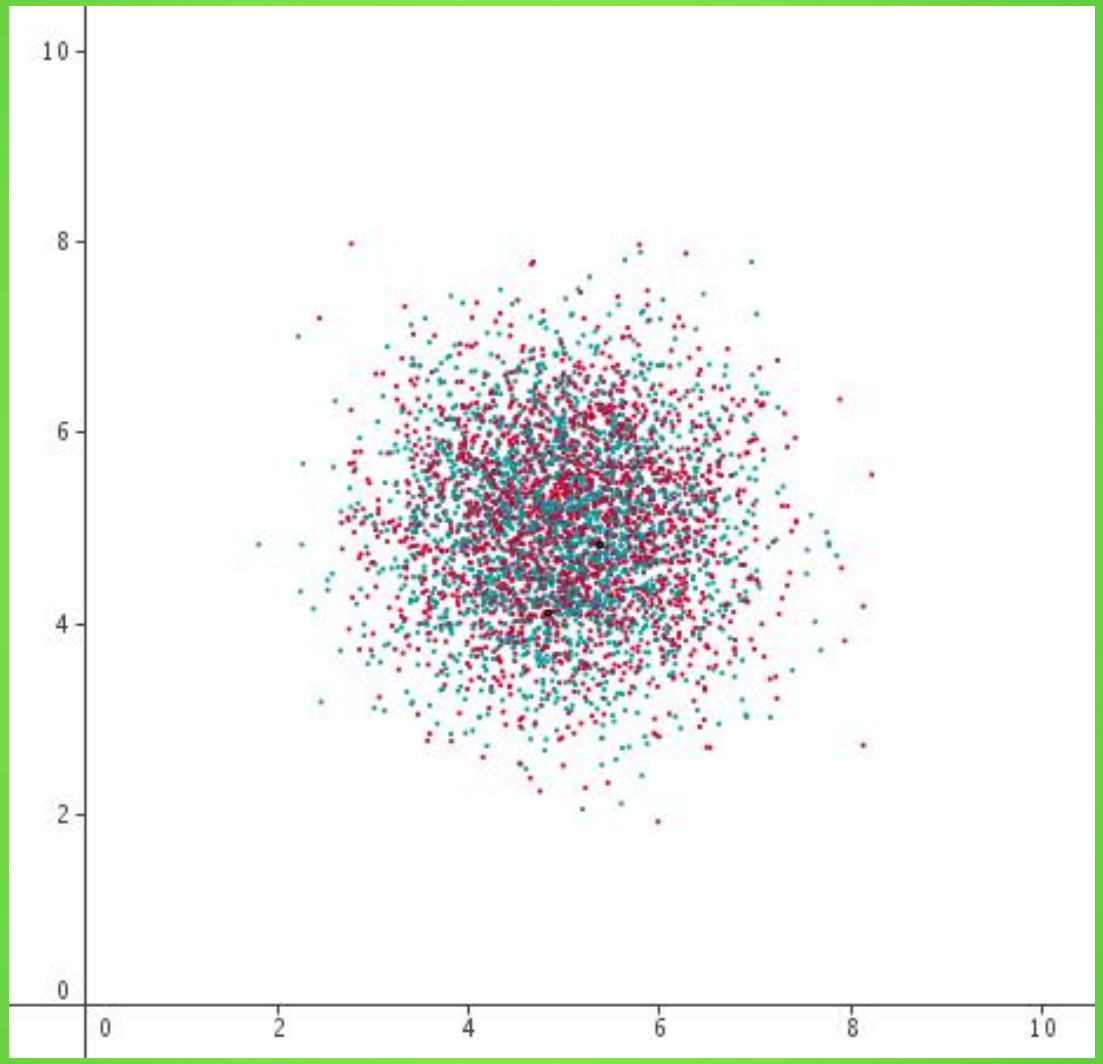
Prediction

* Naive Bayes or Gaussian Classifier
* SVM
* Decision Tree
* Random Forest
* Nearest Neighbors

Prediction


Input data
Nearest Neighbors
Linear SVM
RBF SVM
Gaussian Process
Decision Tree
Random Forest
Neural Net
AdaBoost
Naive Bayes
QDA
.97
.88
.97
.97
.95
.93
.90
.93
.88
.85
.93
.40
.88
.90
.78
.78
.80
.82
.70
.72
.93
.93
.95
.93
.93
.95
.95
.95
.95
.93

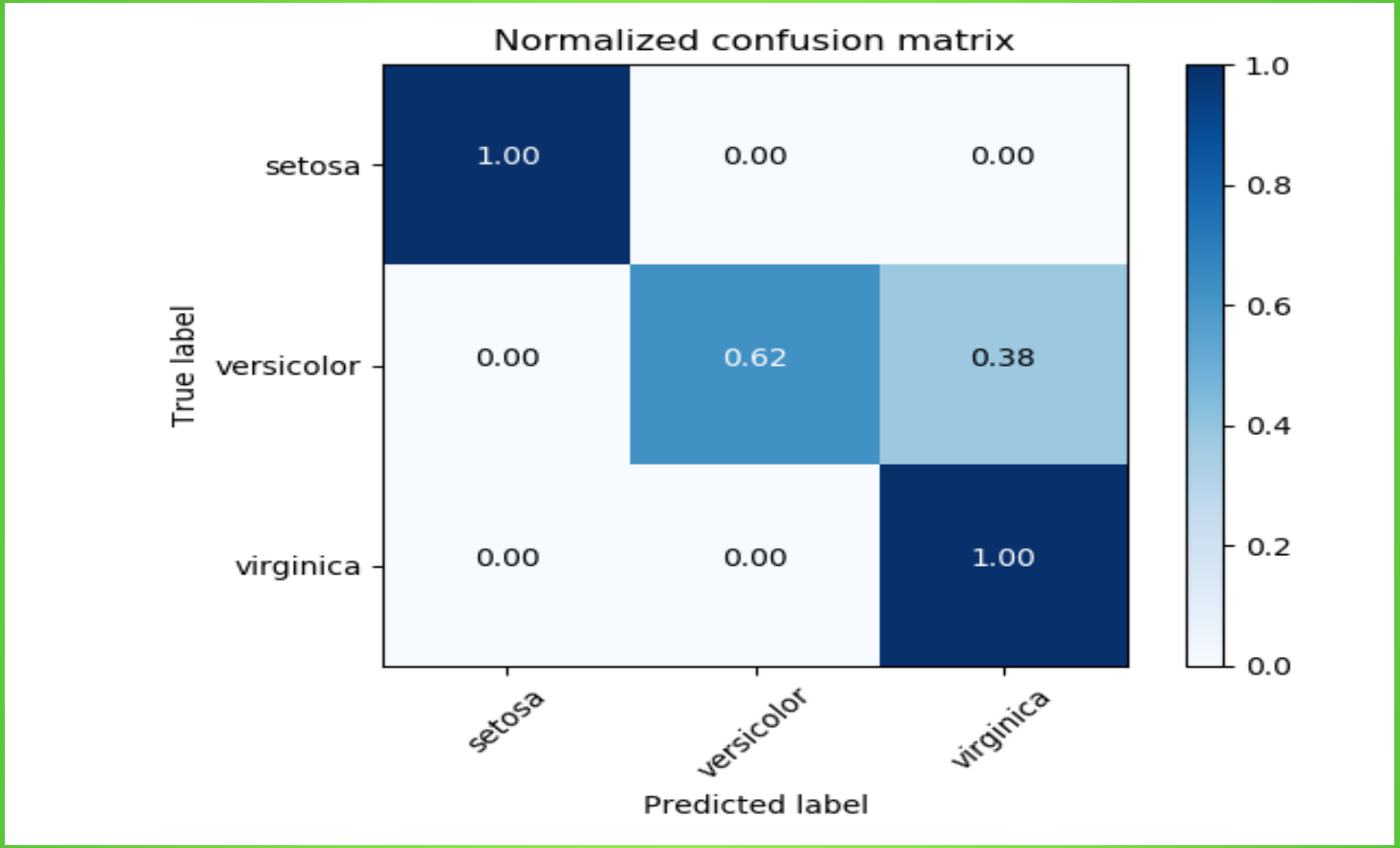
Visualization



Point cloud

Source : http://calque.pagesperso-orange.fr/logiciels/graphisme/nuage.html

Vizualisation



Confusion matrix

Source : https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html

Statistics and preliminary results

Dataset	Num. Examples	Num.Variables /features	Has categorical variables ?	Has missing data ?	Num.Example in each class
Training	5200	1024	no	no	400 per class
Validation	1950	1024	no	no	Unknown
Test	1950	1024	no	no	Unknown

Statistics of the data

Statistics and preliminary results

Method	NaiveBaiyes or Gaussian Classifier	SVM	Decision Tree	Random Forest	Nearest Neighbors
Training	0,81	0,81	0,57	0,64	0,92
CV	0,80	0,79	0,67	0,63	0,83
Validation	0,80	0,80	0,56	0,65	0,85

Preliminary results

Discussion and conclusion