



# MediChal

GROUPE MOSQUITO

# Introduction

- ▶ Paludisme ou Malaria

# Introduction

- ▶ Paludisme ou Malaria
- ▶ Transmission: Moustiques



# Introduction

- ▶ Paludisme ou Malaria
- ▶ Transmission: Moustiques (D'où le nom du groupe)



# Introduction

- ▶ Paludisme ou Malaria
- ▶ Transmission: Moustiques (D'où le nom du groupe)
- ▶ 200 millions infectés/an -> 650 000 morts/an



BUT

RECONNAITRE LES  
CELLULES SAINES DES  
CELLULES PARASITES

# Le Projet

# Le Projet

Que doit-on faire?



# Le Projet

Que doit-on faire?

**Distinguer** des cellules infectés de cellules saines.

# Le Projet

Que doit-on faire?

Distinguer des cellules infectés de cellules saines.

**CLASSIFICATION BINAIRE**

# Le Projet

Que doit-on faire?

Distinguer des cellules infectés de cellules saines.

**CLASSIFICATION BINAIRE**

Ce que l'on a:

# Le Projet

Que doit-on faire?

Distinguer des cellules infectées de cellules saines.

**CLASSIFICATION BINAIRE**

Ce que l'on a:

16 000 images de cellules de personnes atteintes ou non.

# Le Projet

Que doit-on faire?

Distinguer des cellules infectées de cellules saines.

**CLASSIFICATION BINAIRE**

Ce que l'on a:

16 000 images de cellules de personnes atteintes ou non.

On travaille à partir de "features" (ou caractéristiques)

# Le Projet

Que doit-on faire?

Distinguer des cellules infectées de cellules saines.

**CLASSIFICATION BINAIRE**

Ce que l'on a:

16 000 images de cellules de personnes atteintes ou non.

On travaille à partir de "features" (ou caractéristiques)

On en a 19.

# Le Projet

Comment est-on noté?

# Le Projet

Comment est-on noté?

**AUC\_BINARY**



# Le Projet

Comment est-on noté?

**AUC BINARY**

Qu'est-ce que c'est?

Aire sous la courbe de

# Le Projet

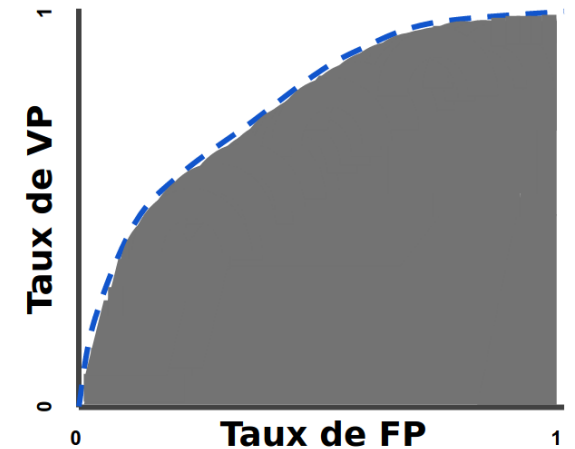
Comment est-on noté?

**AUC BINARY**

Qu'est-ce que c'est?

Aire sous la courbe de:

**Vrais Positifs** en fonction  
des **Faux Positifs**



Dans le domaine médical, il est intéressante de connaître VP et FP...

# Le Projet

Comment est-on noté?

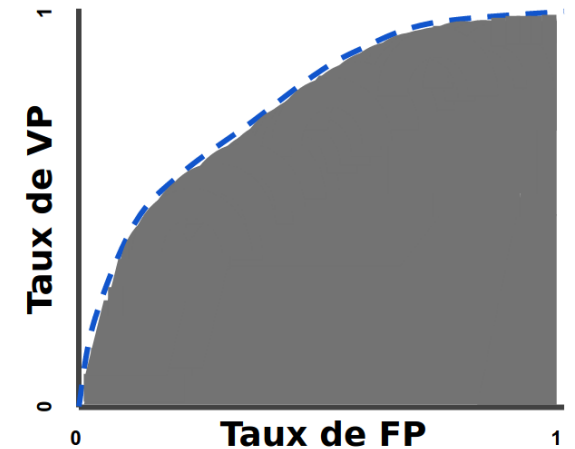
**AUC BINARY**

Qu'est-ce que c'est?

Aire sous la courbe de:

**Vrais Positifs** en fonction  
des **Faux Positifs**

**But: Une aire très proche de 1!**



Dans le domaine médical, il est intéressante de connaître VP et FP...

# Répartition du travail

---

# Répartition du travail

- ▶ 6 personnes -> 3 groupes

# Pre-processing

Tâches:

Pré-traiter les données

Comment?

Avec diverses méthodes et algorithmes

# Pre-processing

Tâche:

Pré-traiter les données

Comment?

Avec diverses méthodes et  
algorithmes

But:

Améliorer les résultats de classification

# Modèle

Tâches:

Entraîner différents classifieurs après pré-  
processing

Choisir le meilleur



# Visualisation

Tâche:

Rendre le problème plus compréhensible

# Visualisation

Tâche:

Rendre le problème plus compréhensible

But:

Comprendre c'est bien

Interprétation des résultats

Amélioration des résultats

Travail effectué

---

# Pre-processing

Sélection de features

# Pre-processing

Sélection de features

Réduction de dimensions

19-->14

Principal Component Analysis

# Pre-processing

Sélection de features

Réduction de dimensions

19-->14

Normalisation

Principal Component Analysis



# Modèle

Modèles testés:

- RandomForest
- AdaBoost
- GaussianNaiveBayes
- KNeighbors
- DecisionTree



# Modèle

Modèles testés:

- ...

Comment déterminer le meilleur modèle?

# Modèle

Modèles testés:

- ...

Comment déterminer le meilleur modèle?

**Validation croisée**

# Modèle

Modèles testés:

- ...

Comment déterminer le meilleur modèle?

Plusieurs ensembles:

- Apprentissage
- Test

**Validation croisée**

# Modèle

Modèles testés:

- ...

Comment déterminer le meilleur modèle?

Plusieurs ensembles:

- ...

Visualisation très importante ici!

**Validation croisée**

# Modèle

Modèles testés:

- ...

Comment déterminer le meilleur modèle?

Plusieurs ensembles:

- ...

Visualisation très importante ici!

Analyse des résultats

**Validation croisée**

# Modèle

Résultat: RandomForest  
Score: 0.9551



Un score assez satisfaisant.



# Conclusion

---

# Conclusion

- ▶ Ce que nous avons appris:
- Collaborer (Communication, Git...)





# Conclusion

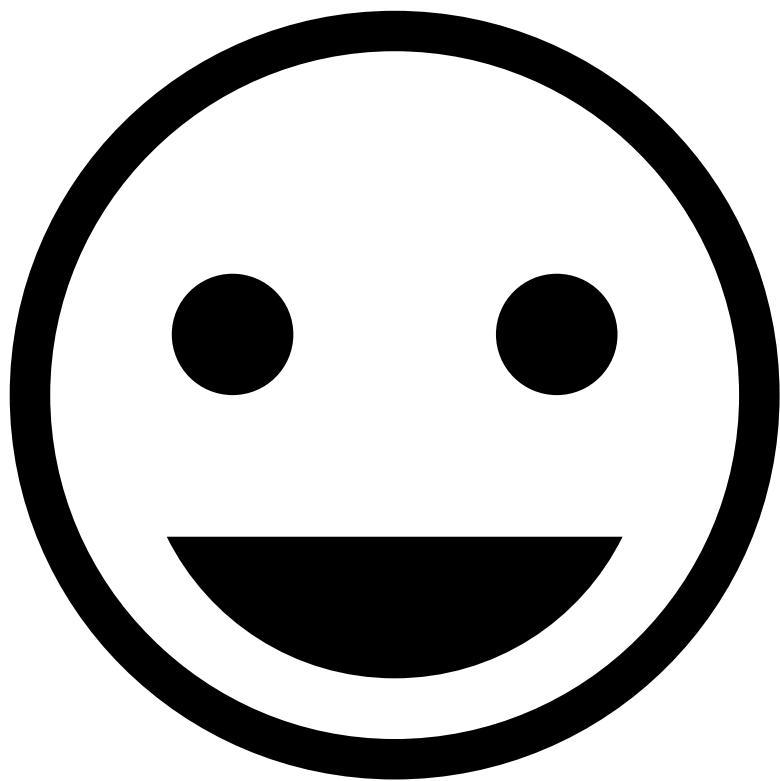
- ▶ Ce que nous avons appris:
- Collaborer (Communication, Git...)
- Base du machine learning et Data Science



# Conclusion

- ▶ Ce que nous avons appris:
  - Collaborer (Communication, Git...)
  - Base du machine learning et Data Science
  - Python et ses librairies
    - Numpy
    - Panda
    - Scikit Learn
    - ...





# Remerciements

ADRIEN PAVAO ET VICTOR ESTRADA