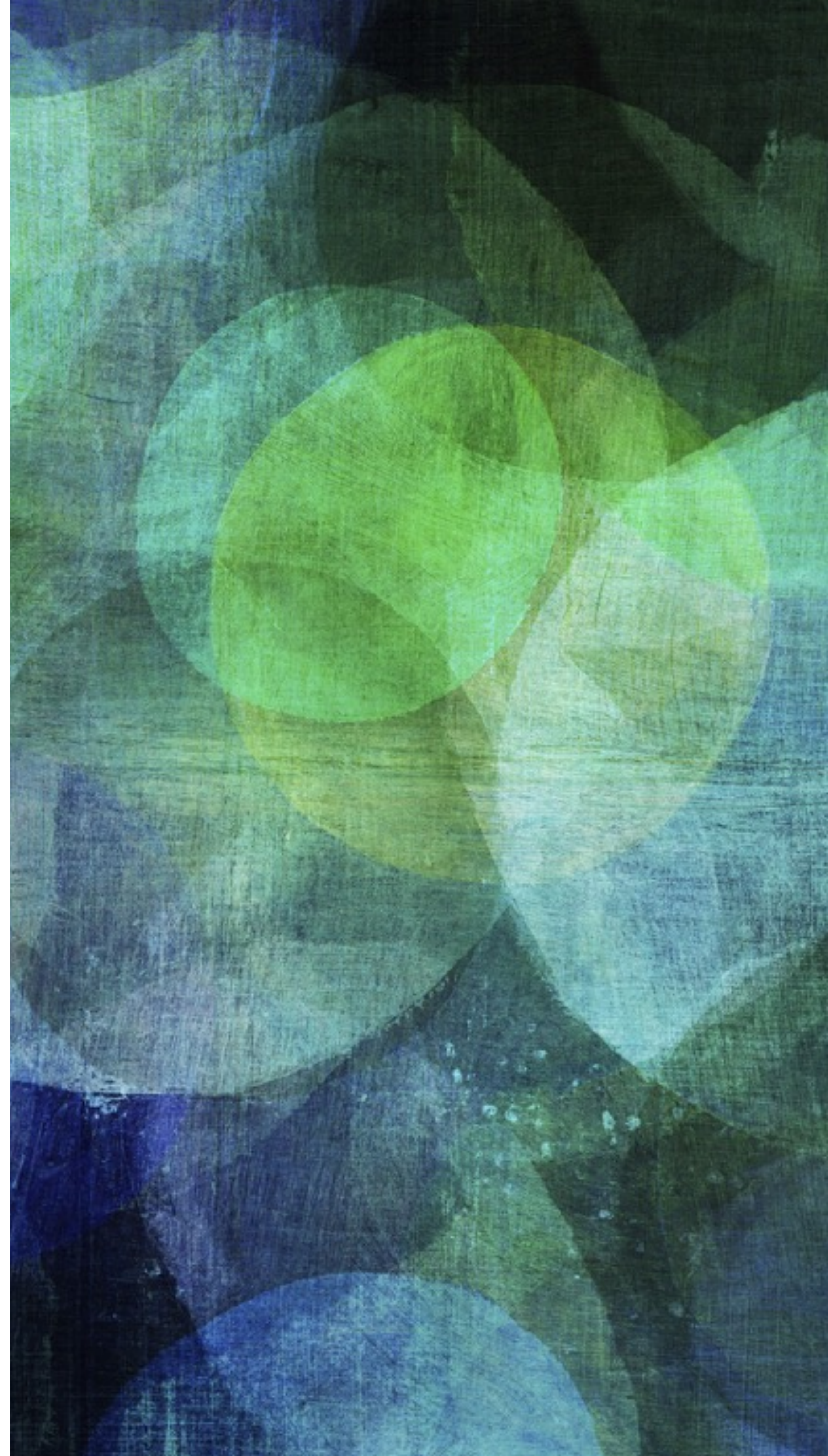
S'ATTAQUER À UNE COMPÉTITION DE MACHINE LEARNING

Méthodologie et exemples pratiques



INTRODUCTION

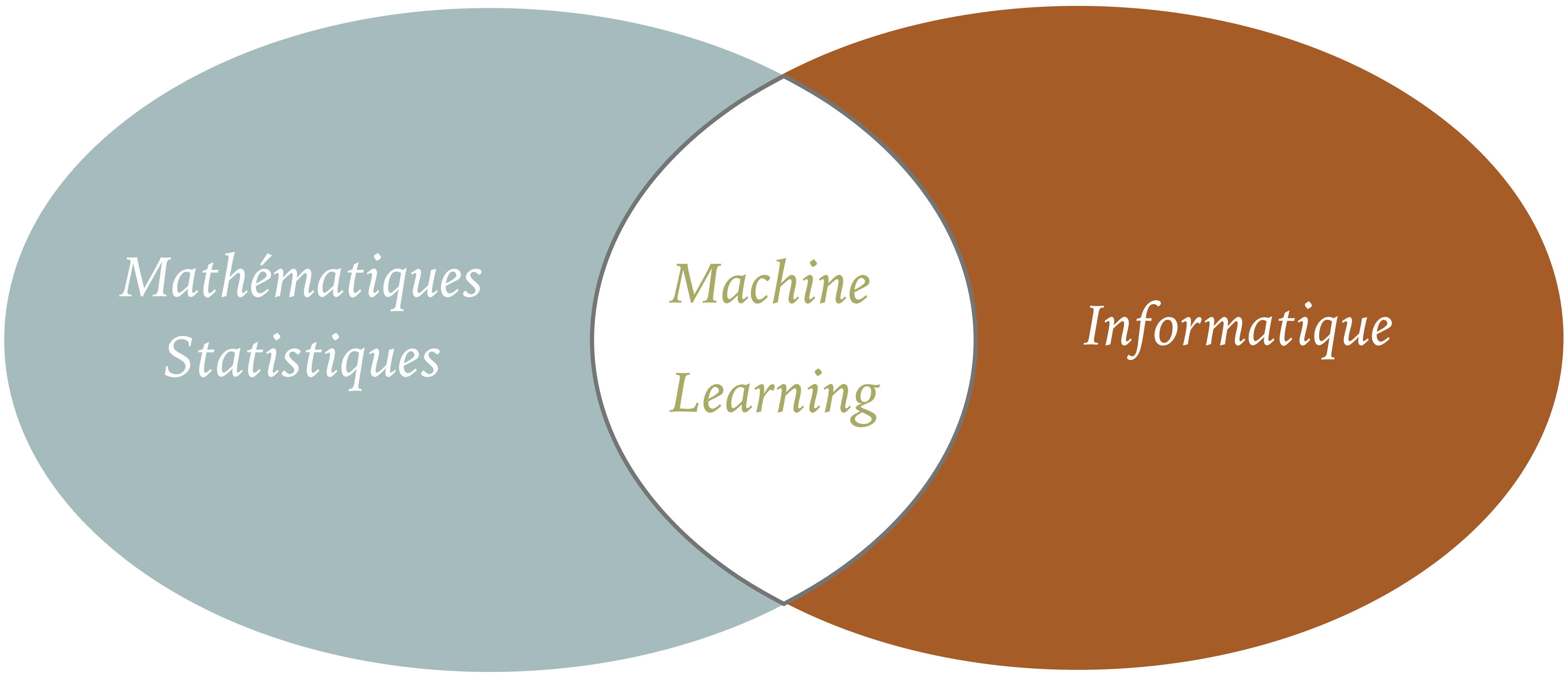
16 Completed Competitions

Competition	Rank
Higgs Boson Machine Learning Challenge Use the ATLAS experiment to identify the Higgs boson Featured · 3 years ago ·	3/1785 Top 1%
Driver Telematics Analysis Use telematic data to identify a driver signature Featured · 3 years ago · tabular data, multiclass classification	3/1528 Top 1%
West Nile Virus Prediction Predict West Nile virus in mosquitos across the city of Chicago Featured · 3 years ago · tabular data, binary classification	3/1306 Top 1%
Truly Native? Predict which web pages served by StumbleUpon are sponsored Featured · 2 years ago · marketing, tabular data, binary classification	3/274 Top 2%
Home Depot Product Search Relevance Predict the relevance of search results on homedepot.com Featured · 2 years ago · tabular data, ranking	7/2125 Top 1%
Data Science Bowl 2017 Can you improve lung cancer detection? Featured · 10 months ago · healthcare, image data, binary classification	10/1972 Top 1%
Caterpillar Tube Pricing Model quoted prices for industrial tube assemblies Featured · 2 years ago · manufacturing, tabular data, regression	12/1323 Top 1%

PLAN

- Qu’est ce qu’une compétition de machine learning ?
- Connaissances requises
- Méthodologie

MACHINE LEARNING



QU'EST CE QU'UNE COMPÉTITION DE MACHINE LEARNING

ANIMAL
CHAT
CHAT
CHIEN
CHIEN

Why the edits made under my username Hardcore Metallica Fan were reverted? They weren't vandalisms, just closure on some GAs after I voted at New York Dolls FAC. And please don't remove the template from the talk page since I'm retired now.

Thanks for uploading Image. I notice the image page specifies that the image is being used under fair use but there is no explanation or rationale as to why it use in Wikipedia article constitutes fair use.

Yes, because the mother of the child in the case against Michael Jackson was studied in here motives and reasonings and judged upon her character just as harshly as Wacko Jacko himself.

Don't tell me to ignore it and incriminate myself. I am going to continue refuting the bullshit that Jayjg keeps throwing at me.

OBSCENE	INSULTE	MENACE
X		X
X		
X	X	
		X

SURFACE	CHAMBRES	PIECES	PRIX
145	3	5	520000
65	1	2	145000
210	4	8	345000
31	1	3	87090

CHANT DES BALEINES
NON
OUI
NON
OUI

QU'EST CE QU'UNE COMPÉTITION DE MACHINE LEARNING

Apprentissage supervisé

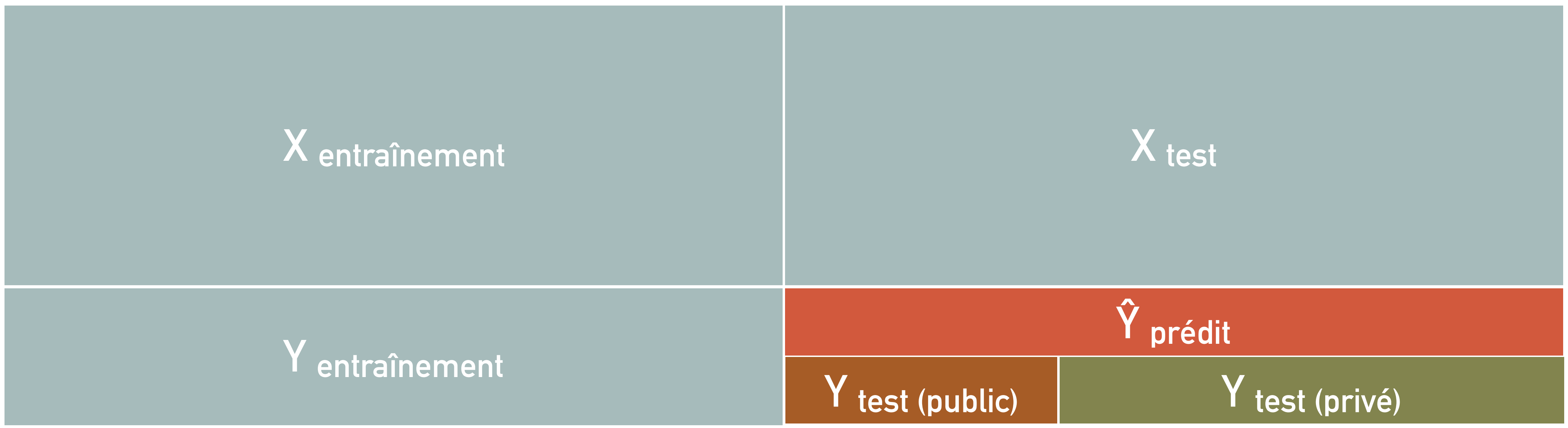
$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ x_{31} & x_{32} & \cdots & x_{3D} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix}$$

- Matrice des variables explicatives

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}$$

- Variable à prédire (cible) contenant les labels (réponses) correspondant à chacune des observations

QU'EST CE QU'UNE COMPÉTITION DE MACHINE LEARNING



Données disponibles pour entraîner votre modèle M

Labels utilisés pour le calcul du score à chaque soumission

Prédiction $M(X_{\text{test}})$

Labels utilisés pour calcul du score final

CONNAISSANCES REQUISES

- Régression linéaire
- Perceptron multicouche
- Arbres décisionnels (RandomForest)
- Boosting (Extreme Gradient Boosting Trees)

- Réseaux de neurones à convolution

- Réseaux de neurones récurrents

CONNAISSANCES REQUISES

- Pourquoi est-ce important ?
  - Parce que c'est passionnant !
  - Savoir les paramétrer efficacement
  - Connaître leurs limites

- Pour mieux les comprendre :
  - Implémenter une version simplifiée
  - Les manipuler sur des données synthétiques

CONNAISSANCES REQUISES

$$\text{Volume\_eau} > \frac{(d \times 0.5 - e)^2 \times 3.14 \times h}{2}$$

CONNAISSANCES REQUISES

- Pourquoi est-ce important ?
  - Parce que c'est passionnant !
  - Savoir les paramétrer efficacement
  - Connaître leurs limites

- Pour mieux les comprendre :
  - Implémenter une version simplifiée
  - Les manipuler sur des données synthétiques

CONNAISSANCES REQUISES

- Langage de programmation
- Bibliothèques utiles (Python)
  - Scikit-learn
  - Keras
  - Pandas
- Structures de données
- Complexité algorithmique
- Gestion de la mémoire

CONNAISSANCES REQUISES

- Pourquoi est-ce important ?
  - Ecrire du code rapidement
  - Ecrire du code optimisé
  - Ecrire du code structuré

CONNAISSANCES REQUISES

- Apprentissage supervisé
  - Régression linéaire
  - Réseau de neurones (MLP)
  - Random forest
  - Boosting
- Apprentissage non-supervisé
  - Analyse en composante principale
  - K-Means
- Notions
  - Sur-apprentissage/Sous-apprentissage
  - Variables numériques/Catégorielles
  - Validation croisée

- Bibliothèques utiles (Python)
  - Scikit-learn
  - Keras
  - Pandas
  - Numpy
- Structure de données
  - Table de hachage
  - Arbre binaire
  - Liste et vecteur
- Algorithmique
  - Calcul de complexité
  - Méthode de tri

METHODOLOGIE

- Passage en revue de l'état de l'art (5 %)
- Exploration des données (20 %)
- Features engineering (50 %)
- Mise en place d'une procédure d'évaluation (10 %)
- Création du modèle (15 %)

METHODOLOGIE : EXPLORATION DES DONNEES

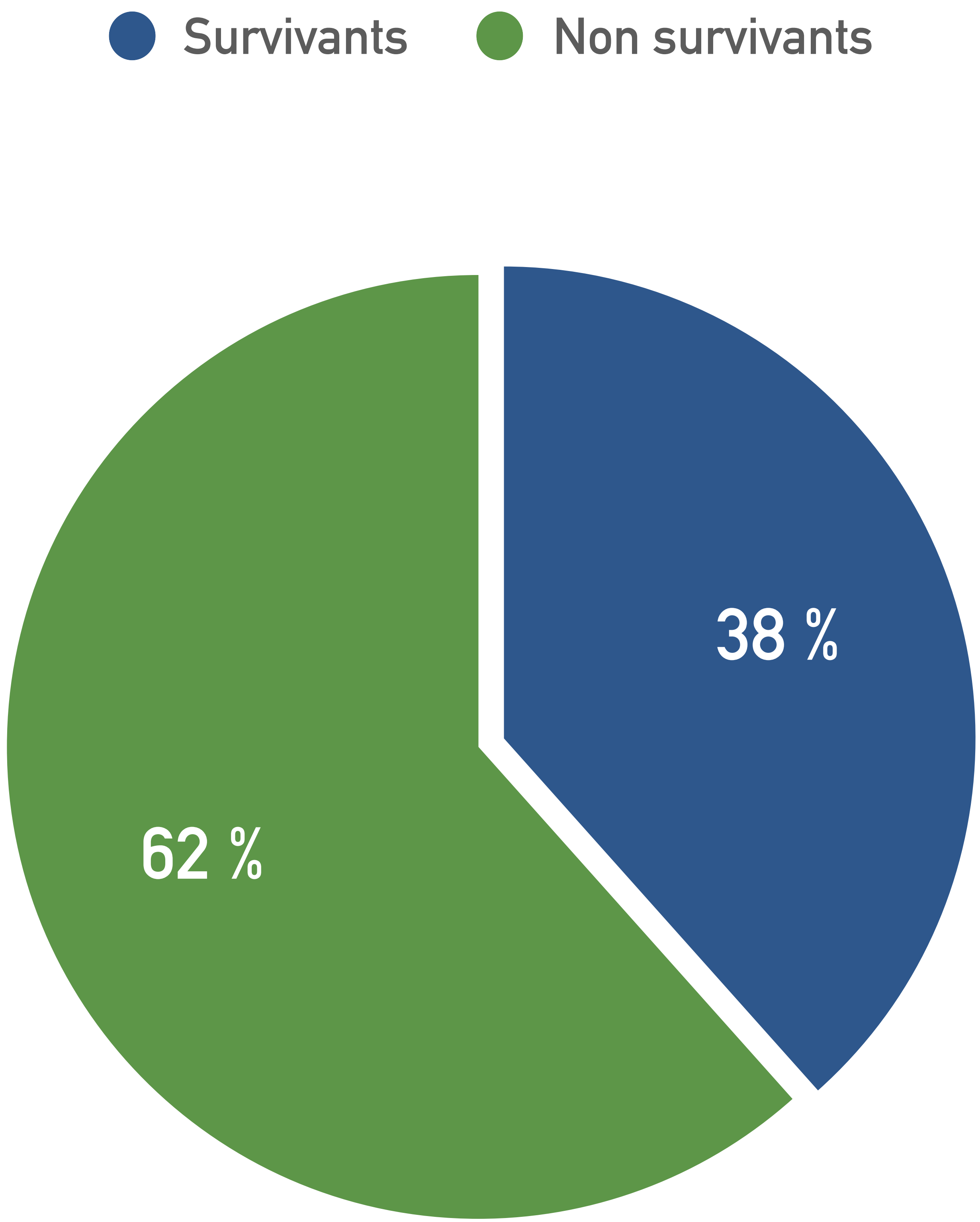
- Compréhension de ce que représente chaque variable
- Analyse de la variable cible (données équilibrées ?)
- Type des variables explicatives : Catégorielles, Ordinales, Continues
- Analyse des variables explicatives
- Valeurs manquantes
- Corrélation entre les variables

- **Se familiariser avec le jeu de données**

METHODOLOGIE : EXPLORATION DES DONNEES

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
276	1	Upper	Andrews, Miss. Kornelia Theodosia	female	63	1	0	13502	77.9583	D7	S
277	0	Lower	Lindblom, Miss. Augusta Charlotta	female	45	0	0	347073	7.75	nan	S
97	0	Upper	Goldschmidt, Mr. George B	male	71	0	0	PC 17754	34.6542	A5	C
837	0	Lower	Pasic, Mr. Jakob	male	21	0	0	315097	8.6625	nan	S
388	1	Middle	Buss, Miss. Kate	female	36	0	0	27849	13	nan	S
23	1	Lower	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292	nan	Q
177	0	Lower	Lefebre, Master. Henry Forbes	male	nan	3	1	4133	25.4667	nan	S
233	0	Middle	Sjostedt, Mr. Ernst Adolf	male	59	0	0	237442	13.5	nan	S
376	1	Upper	Meyer, Mrs. Edgar Joseph (Leila Saks)	female	nan	1	0	PC 17604	82.1708	nan	C
402	0	Lower	Adams, Mr. John	male	26	0	0	341826	8.05	nan	S
627	0	Middle	Kirkland, Rev. Charles Leonard	male	57	0	0	219533	12.35	nan	Q
435	0	Upper	Silvey, Mr. William Baird	male	50	1	0	13507	55.9	E44	S
660	0	Upper	Newell, Mr. Arthur Webster	male	58	0	2	35273	113.275	D48	C
288	0	Lower	Naidenoff, Mr. Penko	male	22	0	0	349206	7.8958	nan	S
101	0	Lower	Petranec, Miss. Matilda	female	28	0	0	349245	7.8958	nan	S

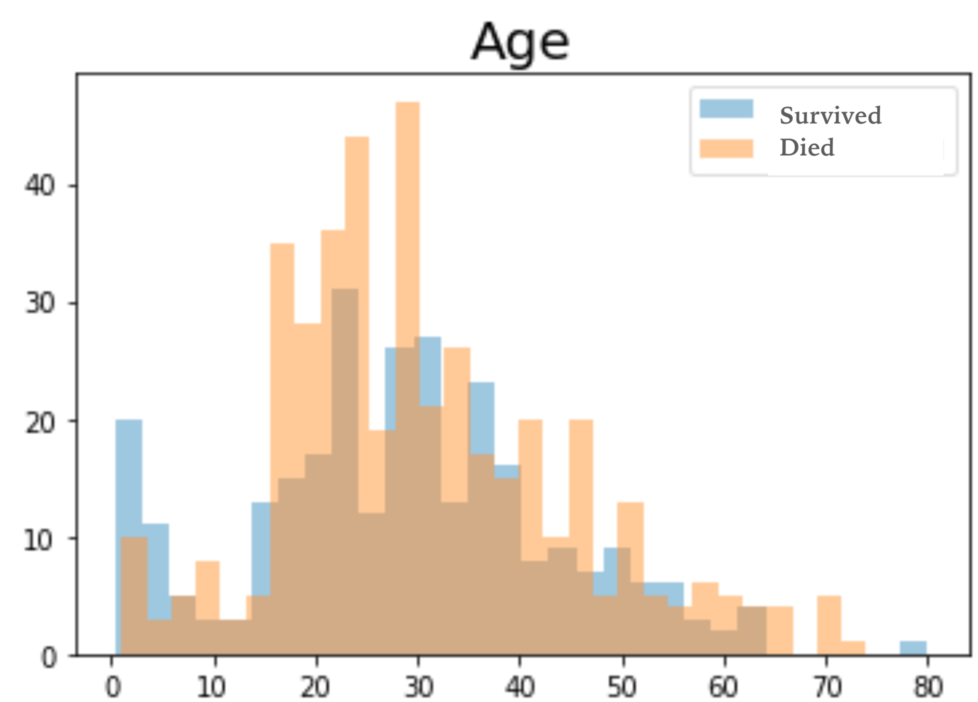
METHODOLOGIE : EXPLORATION DES DONNEES

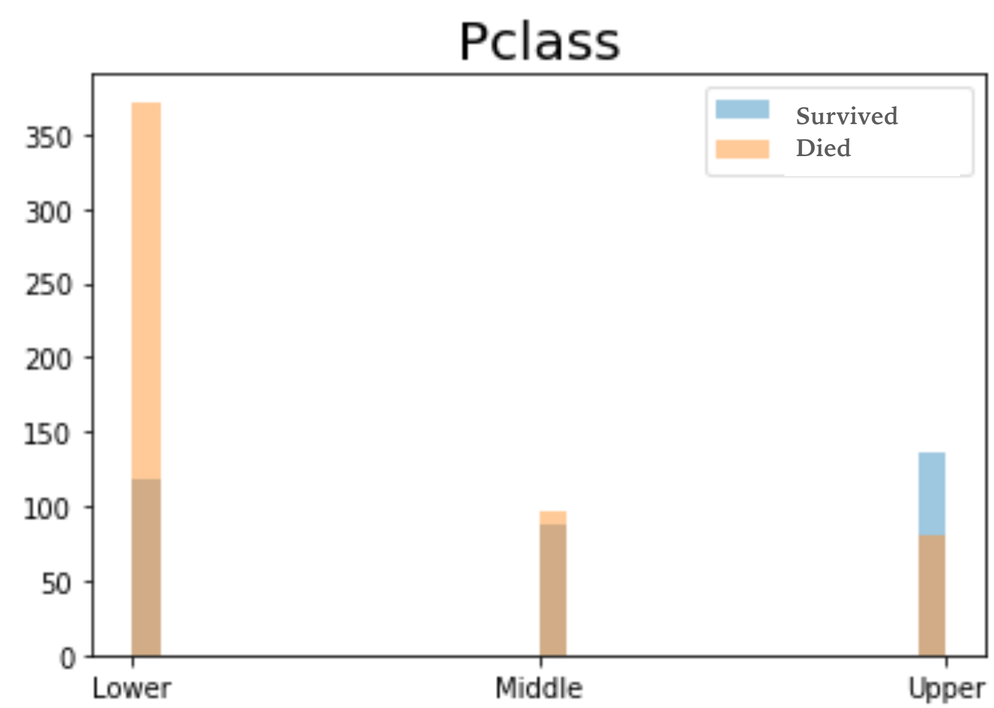


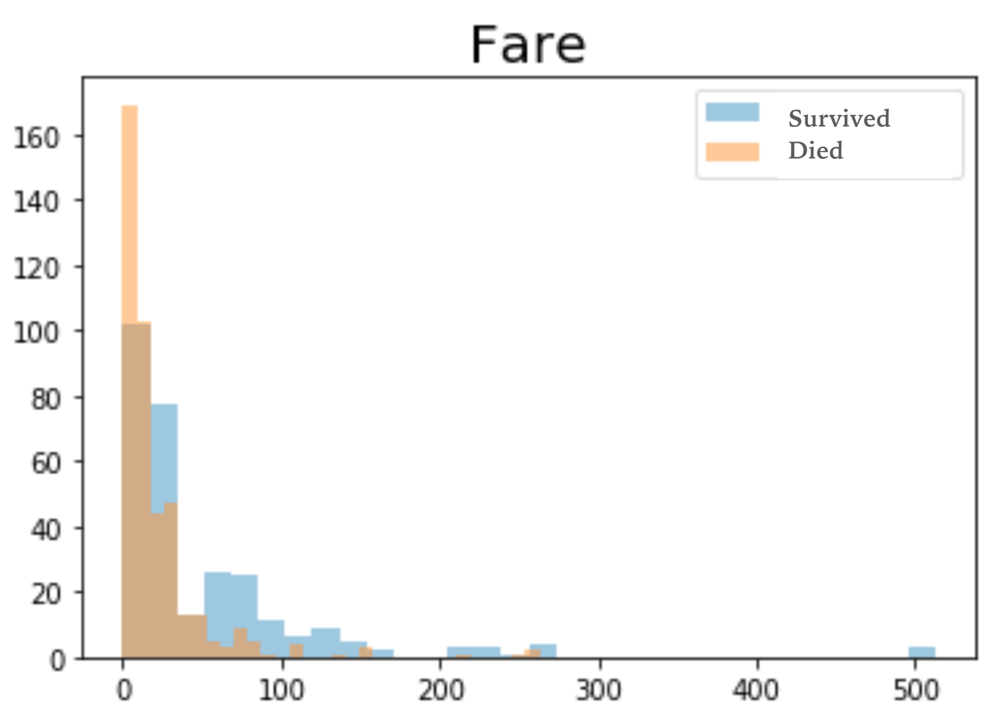
Valeurs manquantes

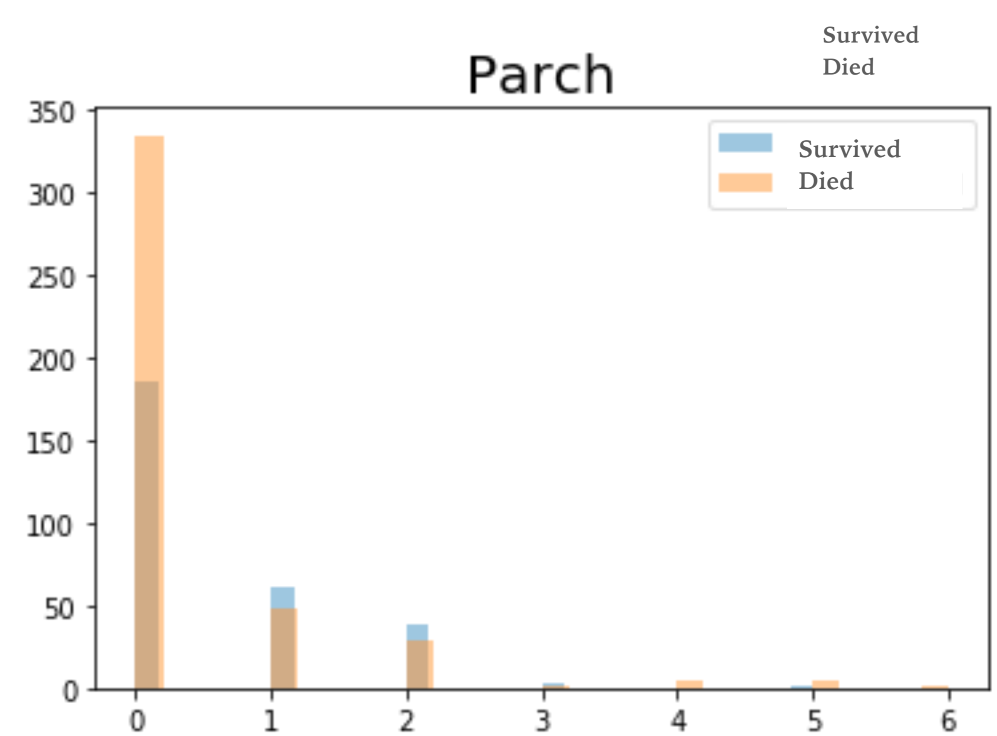
	count	ratio
PassengerId	0	0
Survived	0	0
Pclass	0	0
Name	0	0
Sex	0	0
Age	177	0.198653
SibSp	0	0
Parch	0	0
Ticket	0	0
Fare	0	0
Cabin	687	0.771044
Embarked	2	0.00224467

METHODOLOGIE : EXPLORATION DES DONNEES

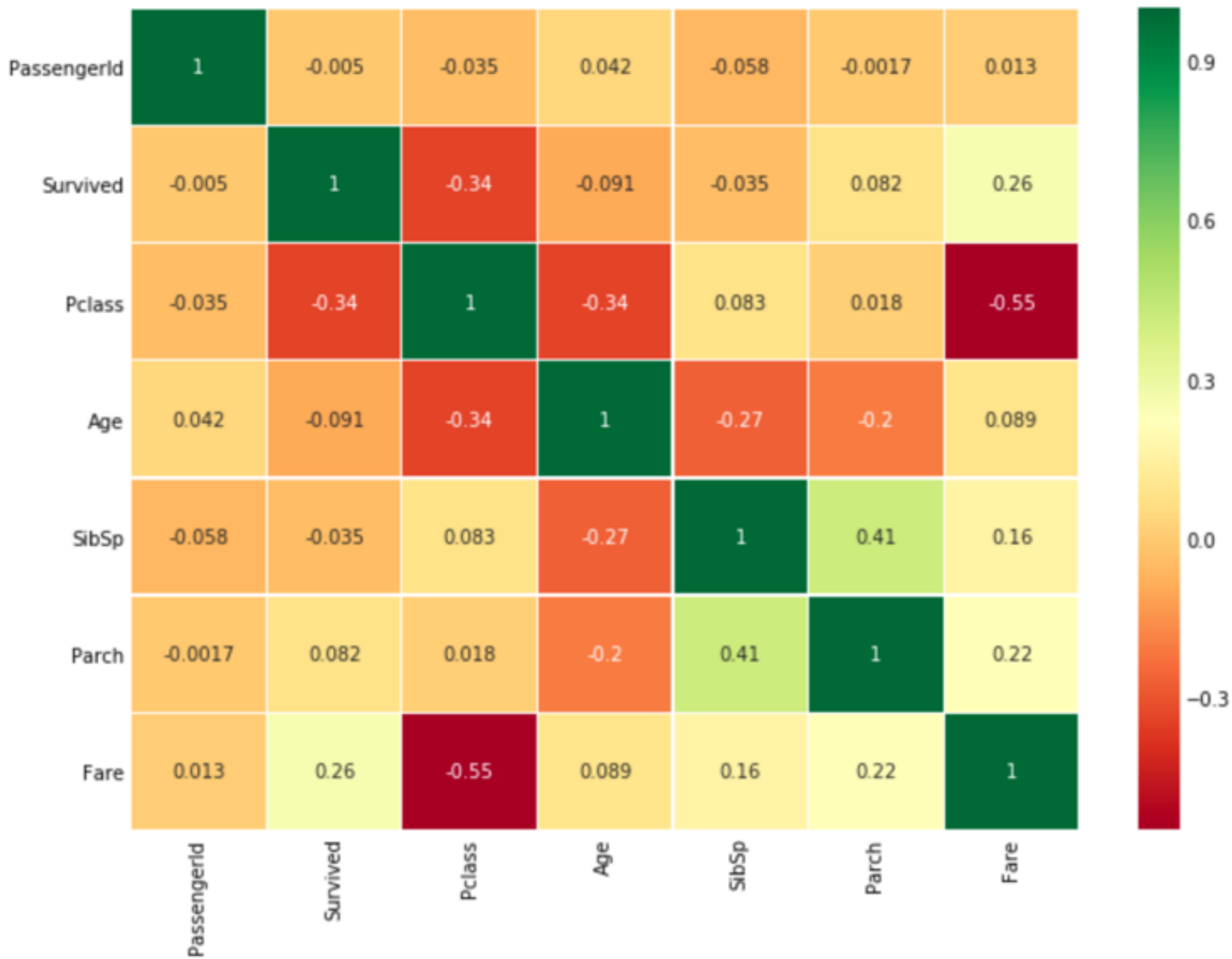








METHODOLOGIE : EXPLORATION DES DONNEES



METHODOLOGIE : FEATURE ENGINEERING

- Transformation de la donnée pour être exploitée au mieux par les algorithmes d'apprentissage
- Extraction des informations pertinentes et informatives cachées dans les données
- Sélection de variables explicatives

*« Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work. Feature engineering is fundamental to the application of machine learning, and is both difficult and expensive. Coming up with features is difficult, time-consuming, requires expert knowledge. "**Applied machine learning**" **is basically feature engineering**. »*

Andrew Ng

METHODOLOGIE : FEATURE ENGINEERING

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
465	0	Lower	Maisner, Mr. Simon	male	nan	0	0	A/S 2816	8.05	nan	S
84	0	Upper	Carrau, Mr. Francisco M	male	28	0	0	113059	47.1	nan	S
780	1	Upper	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)	female	43	0	1	24160	211.338	B3	S
852	0	Lower	Svensson, Mr. Johan	male	74	0	0	347060	7.775	nan	S
748	1	Middle	Sinkkonen, Miss. Anna	female	30	0	0	250648	13	nan	S
362	0	Middle	del Carlo, Mr. Sebastiano	male	29	1	0	SC/PARIS 2167	27.7208	nan	C
200	0	Middle	Yrois, Miss. Henriette ("Mrs Harbeck")	female	24	0	0	248747	13	nan	S
630	0	Lower	O'Connell, Mr. Patrick D	male	nan	0	0	334912	7.7333	nan	Q
732	0	Lower	Hassan, Mr. Houssein G N	male	11	0	0	2699	18.7875	nan	C
175	0	Upper	Smith, Mr. James Clinch	male	56	0	0	17764	30.6958	A7	C
538	1	Upper	LeRoy, Miss. Bertha	female	30	0	0	PC 17761	106.425	nan	C
384	1	Upper	Holverson, Mrs. Alexander Oskar (Mary Aline Towner)	female	35	1	0	113789	52	nan	S
881	1	Middle	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433	26	nan	S
173	1	Lower	Johnson, Miss. Eleanor Ileen	female	1	1	1	347742	11.1333	nan	S
327	0	Lower	Nysveen, Mr. Johan Hansen	male	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
465	0	Lower	Maisner, Mr. Simon	male	nan	0	0	A/S 2816	8.05	nan	S
84	0	Upper	Carrau, Mr. Francisco M	male	28	0	0	113059	47.1	nan	S
780	1	Upper	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)	female	43	0	1	24160	211.338	B3	S
852	0	Lower	Svensson, Mr. Johan	male	74	0	0	347060	7.775	nan	S
748	1	Middle	Sinkkonen, Miss. Anna	female	30	0	0	250648	13	nan	S
362	0	Middle	del Carlo, Mr. Sebastiano	male	29	1	0	SC/PARIS 2167	27.7208	nan	C
200	0	Middle	Yrois, Miss. Henriette ("Mrs Harbeck")	female	24	0	0	248747	13	nan	S
630	0	Lower	O'Connell, Mr. Patrick D	male	nan	0	0	334912	7.7333	nan	Q
732	0	Lower	Hassan, Mr. Houssein G N	male	11	0	0	2699	18.7875	nan	C
175	0	Upper	Smith, Mr. James Clinch	male	56	0	0	17764	30.6958	A7	C
538	1	Upper	LeRoy, Miss. Bertha	female	30	0	0	PC 17761	106.425	nan	C
384	1	Upper	Holverson, Mrs. Alexander Oskar (Mary Aline Towner)	female	35	1	0	113789	52	nan	S
881	1	Middle	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433	26	nan	S
173	1	Lower	Johnson, Miss. Eleanor Ileen	female	1	1	1	347742	11.1333	nan	S
327	0	Lower	Nysveen, Mr. Johan Hansen	male	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	Maisner, Mr. Simon	male	nan	0	0	A/S 2816	8.05	nan	S
0	1	Carrau, Mr. Francisco M	male	28	0	0	113059	47.1	nan	S
1	1	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)	female	43	0	1	24160	211.338	B3	S
0	3	Svensson, Mr. Johan	male	74	0	0	347060	7.775	nan	S
1	2	Sinkkonen, Miss. Anna	female	30	0	0	250648	13	nan	S
0	2	del Carlo, Mr. Sebastiano	male	29	1	0	SC/PARIS 2167	27.7208	nan	C
0	2	Yrois, Miss. Henriette ("Mrs Harbeck")	female	24	0	0	248747	13	nan	S
0	3	O'Connell, Mr. Patrick D	male	nan	0	0	334912	7.7333	nan	Q
0	3	Hassan, Mr. Houssein G N	male	11	0	0	2699	18.7875	nan	C
0	1	Smith, Mr. James Clinch	male	56	0	0	17764	30.6958	A7	C
1	1	LeRoy, Miss. Bertha	female	30	0	0	PC 17761	106.425	nan	C
1	1	Holverson, Mrs. Alexander Oskar (Mary Aline Towner)	female	35	1	0	113789	52	nan	S
1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433	26	nan	S
1	3	Johnson, Miss. Eleanor Ileen	female	1	1	1	347742	11.1333	nan	S
0	3	Nysveen, Mr. Johan Hansen	male	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	True	False	False	male	nan	0	0	A/S 2816	8.05	nan	S
0	1	True	False	False	male	28	0	0	113059	47.1	nan	S
1	1	True	True	False	female	43	0	1	24160	211.338	B3	S
0	3	True	False	False	male	74	0	0	347060	7.775	nan	S
1	2	False	False	True	female	30	0	0	250648	13	nan	S
0	2	True	False	False	male	29	1	0	SC/PARIS 2167	27.7208	nan	C
0	2	True	True	True	female	24	0	0	248747	13	nan	S
0	3	True	False	False	male	nan	0	0	334912	7.7333	nan	Q
0	3	True	False	False	male	11	0	0	2699	18.7875	nan	C
0	1	True	False	False	male	56	0	0	17764	30.6958	A7	C
1	1	False	False	True	female	30	0	0	PC 17761	106.425	nan	C
1	1	True	True	False	female	35	1	0	113789	52	nan	S
1	2	True	True	False	female	25	0	1	230433	26	nan	S
1	3	False	False	True	female	1	1	1	347742	11.1333	nan	S
0	3	True	False	False	male	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	True	False	False	0	1	nan	0	0	A/S 2816	8.05	nan	S
0	1	True	False	False	0	1	28	0	0	113059	47.1	nan	S
1	1	True	True	False	1	0	43	0	1	24160	211.338	B3	S
0	3	True	False	False	0	1	74	0	0	347060	7.775	nan	S
1	2	False	False	True	1	0	30	0	0	250648	13	nan	S
0	2	True	False	False	0	1	29	1	0	SC/PARIS 2167	27.7208	nan	C
0	2	True	True	True	1	0	24	0	0	248747	13	nan	S
0	3	True	False	False	0	1	nan	0	0	334912	7.7333	nan	Q
0	3	True	False	False	0	1	11	0	0	2699	18.7875	nan	C
0	1	True	False	False	0	1	56	0	0	17764	30.6958	A7	C
1	1	False	False	True	1	0	30	0	0	PC 17761	106.425	nan	C
1	1	True	True	False	1	0	35	1	0	113789	52	nan	S
1	2	True	True	False	1	0	25	0	1	230433	26	nan	S
1	3	False	False	True	1	0	1	1	1	347742	11.1333	nan	S
0	3	True	False	False	0	1	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	True	False	False	0	1	30	0	0	A/S 2816	8.05	nan	S
0	1	True	False	False	0	1	28	0	0	113059	47.1	nan	S
1	1	True	True	False	1	0	43	0	1	24160	211.338	B3	S
0	3	True	False	False	0	1	74	0	0	347060	7.775	nan	S
1	2	False	False	True	1	0	30	0	0	250648	13	nan	S
0	2	True	False	False	0	1	29	1	0	SC/PARIS 2167	27.7208	nan	C
0	2	True	True	True	1	0	24	0	0	248747	13	nan	S
0	3	True	False	False	0	1	30	0	0	334912	7.7333	nan	Q
0	3	True	False	False	0	1	11	0	0	2699	18.7875	nan	C
0	1	True	False	False	0	1	56	0	0	17764	30.6958	A7	C
1	1	False	False	True	1	0	30	0	0	PC 17761	106.425	nan	C
1	1	True	True	False	1	0	35	1	0	113789	52	nan	S
1	2	True	True	False	1	0	25	0	1	230433	26	nan	S
1	3	False	False	True	1	0	1	1	1	347742	11.1333	nan	S
0	3	True	False	False	0	1	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	3	True	False	False	0	1	30	0	0	A/S 2816	8.05	nan	S
0	1	True	False	False	0	1	28	0	0	113059	47.1	nan	S
1	1	True	True	False	1	0	43	0	1	24160	211.338	B3	S
0	3	True	False	False	0	1	74	0	0	347060	7.775	nan	S
1	2	False	False	True	1	0	30	0	0	250648	13	nan	S
0	2	True	False	False	0	1	29	1	0	SC/PARIS 2167	27.7208	nan	C
0	2	True	True	True	1	0	24	0	0	248747	13	nan	S
0	3	True	False	False	0	1	30	0	0	334912	7.7333	nan	Q
0	3	True	False	False	0	1	11	0	0	2699	18.7875	nan	C
0	1	True	False	False	0	1	56	0	0	17764	30.6958	A7	C
1	1	False	False	True	1	0	30	0	0	PC 17761	106.425	nan	C
1	1	True	True	False	1	0	35	1	0	113789	52	nan	S
1	2	True	True	False	1	0	25	0	1	230433	26	nan	S
1	3	False	False	True	1	0	1	1	1	347742	11.1333	nan	S
0	3	True	False	False	0	1	61	0	0	345364	6.2375	nan	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	Fare	Embarked
0	3	True	False	False	0	1	30	0	0	8.05	S
0	1	True	False	False	0	1	28	0	0	47.1	S
1	1	True	True	False	1	0	43	0	1	211.338	S
0	3	True	False	False	0	1	74	0	0	7.775	S
1	2	False	False	True	1	0	30	0	0	13	S
0	2	True	False	False	0	1	29	1	0	27.7208	C
0	2	True	True	True	1	0	24	0	0	13	S
0	3	True	False	False	0	1	30	0	0	7.7333	Q
0	3	True	False	False	0	1	11	0	0	18.7875	C
0	1	True	False	False	0	1	56	0	0	30.6958	C
1	1	False	False	True	1	0	30	0	0	106.425	C
1	1	True	True	False	1	0	35	1	0	52	S
1	2	True	True	False	1	0	25	0	1	26	S
1	3	False	False	True	1	0	1	1	1	11.1333	S
0	3	True	False	False	0	1	61	0	0	6.2375	S

METHODOLOGIE : FEATURE ENGINEERING

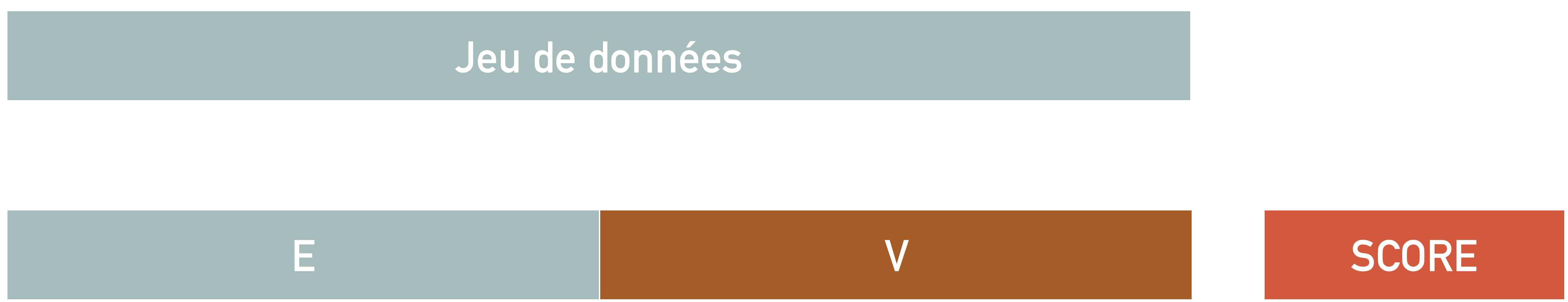
Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	familySize	Fare	Embarked
0	3	True	False	False	0	1	30	0	0	1	8.05	S
0	1	True	False	False	0	1	28	0	0	1	47.1	S
1	1	True	True	False	1	0	43	0	1	2	211.338	S
0	3	True	False	False	0	1	74	0	0	1	7.775	S
1	2	False	False	True	1	0	30	0	0	1	13	S
0	2	True	False	False	0	1	29	1	0	2	27.7208	C
0	2	True	True	True	1	0	24	0	0	1	13	S
0	3	True	False	False	0	1	30	0	0	1	7.7333	Q
0	3	True	False	False	0	1	11	0	0	1	18.7875	C
0	1	True	False	False	0	1	56	0	0	1	30.6958	C
1	1	False	False	True	1	0	30	0	0	1	106.425	C
1	1	True	True	False	1	0	35	1	0	2	52	S
1	2	True	True	False	1	0	25	0	1	2	26	S
1	3	False	False	True	1	0	1	1	1	3	11.1333	S
0	3	True	False	False	0	1	61	0	0	1	6.2375	S

METHODOLOGIE : FEATURE ENGINEERING

Survived	Pclass	Name_Mr	Name_Mrs	Name_Miss	Sex_female	Sex_male	Age	SibSp	Parch	Fare	familySize	Emb_C	Emb_Q	Emb_S
0	3	True	False	False	0	1	30	0	0	8.05	1	0	0	1
0	1	True	False	False	0	1	28	0	0	47.1	1	0	0	1
1	1	True	True	False	1	0	43	0	1	211.338	2	0	0	1
0	3	True	False	False	0	1	74	0	0	7.775	1	0	0	1
1	2	False	False	True	1	0	30	0	0	13	1	0	0	1
0	2	True	False	False	0	1	29	1	0	27.7208	2	1	0	0
0	2	True	True	True	1	0	24	0	0	13	1	0	0	1
0	3	True	False	False	0	1	30	0	0	7.7333	1	0	1	0
0	3	True	False	False	0	1	11	0	0	18.7875	1	1	0	0
0	1	True	False	False	0	1	56	0	0	30.6958	1	1	0	0
1	1	False	False	True	1	0	30	0	0	106.425	1	1	0	0
1	1	True	True	False	1	0	35	1	0	52	2	0	0	1
1	2	True	True	False	1	0	25	0	1	26	2	0	0	1
1	3	False	False	True	1	0	1	1	1	11.1333	3	0	0	1
0	3	True	False	False	0	1	61	0	0	6.2375	1	0	0	1

METHODOLOGIE : PROCEDURE D'EVALUATION

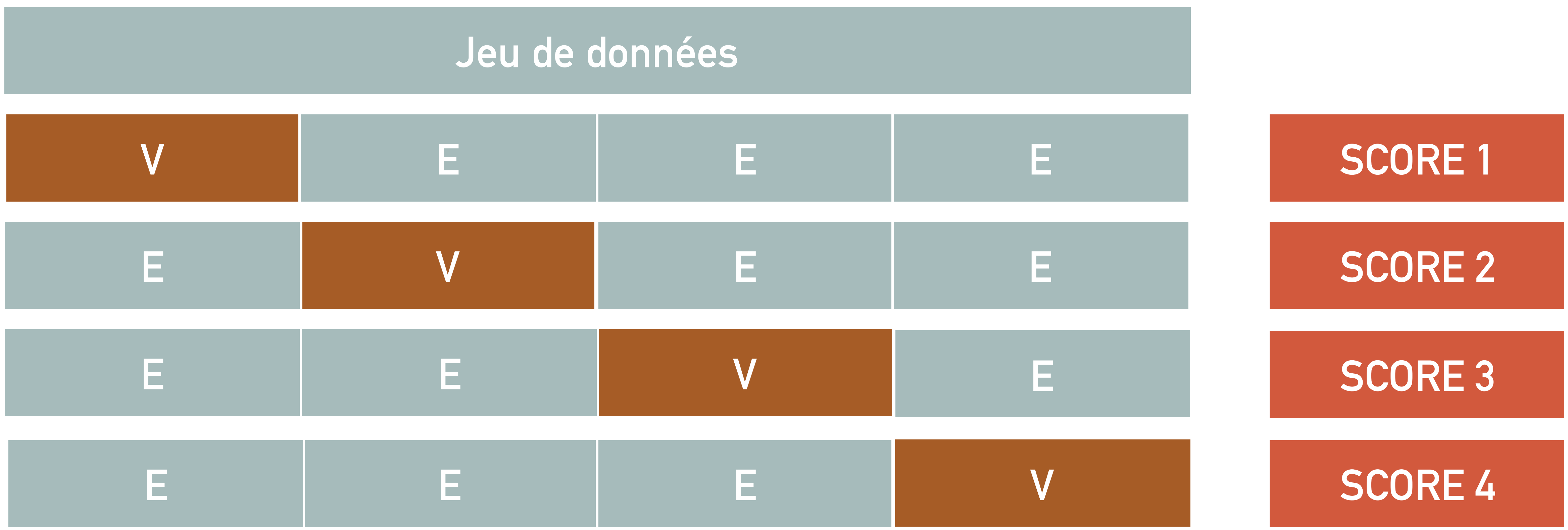
- Découpage aléatoire entraînement / validation (50% / 50%)



- Entraînement sur E puis évaluation du modèle sur V
- Inconvénients :
  - Surévaluation ou sous-évaluation du modèle
  - Utilisation d'un sous-ensemble des données pour l'entraînement

METHODOLOGIE : PROCEDURE D’EVALUATION

- Validation croisée : découpage en k parties (k-fold)



- Pour chacune des k parties, validation sur k et entrainement du modèle sur les k-1 parties restantes

METHODOLOGIE : CREATION DES MODELES

- Entraînement et évaluation d'un modèle :

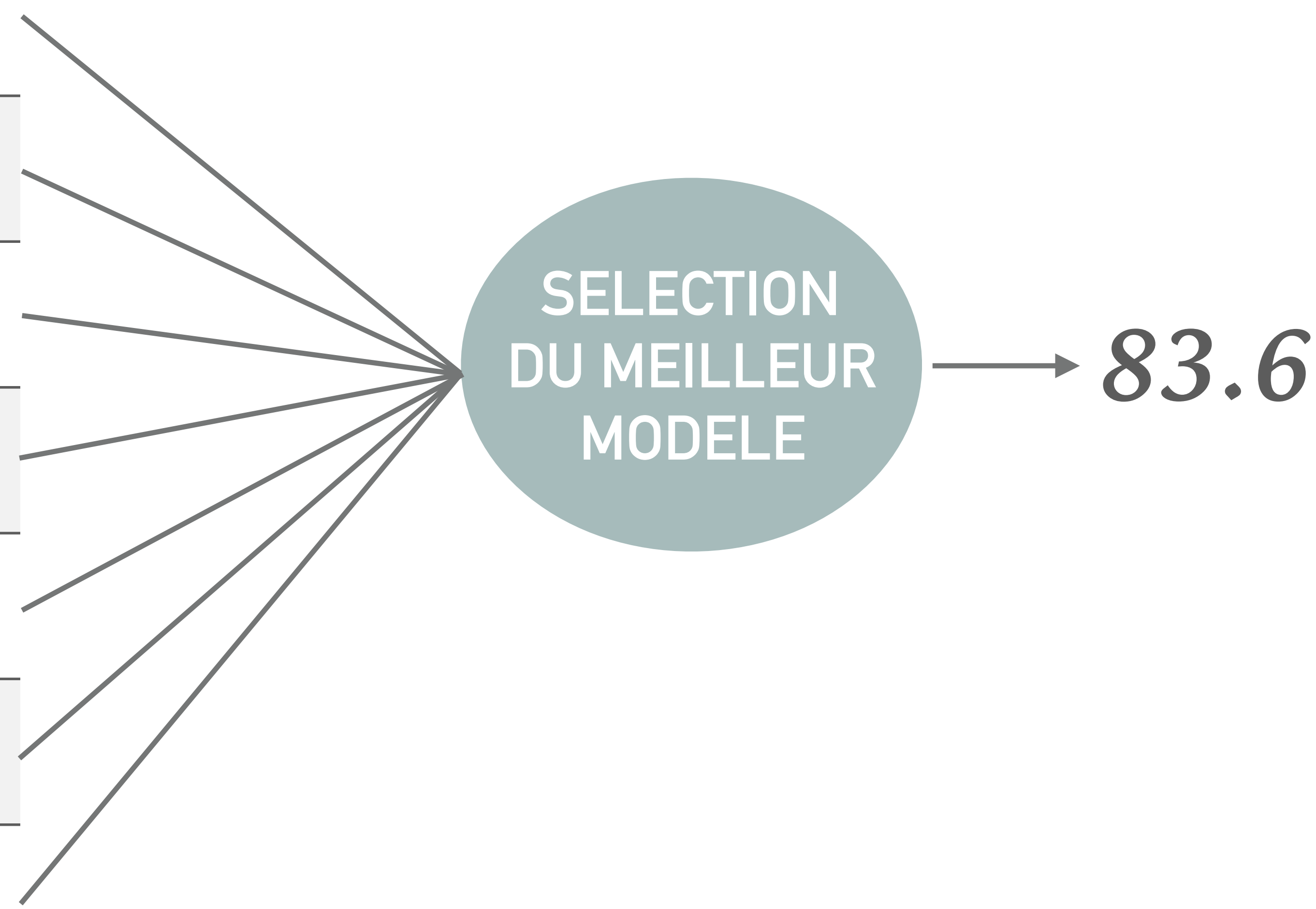
```
>>> from sklearn.model_selection import cross_val_score
>>> clf = svm.SVC(kernel='linear', C=1)
>>> scores = cross_val_score(clf, iris.data, iris.target, cv=5)
>>> scores
array([ 0.96...,  1.  ...,  0.96...,  0.96...,  1.        ])
```

- Ajuster les hyper-paramètres :
  - Manuellement
  - Automatiquement
    - Hyperopt
    - Grid search

Modèle	Score en VC
RandomForest	82.3
ExtraTrees	82.7
Xgboost	83.6
k-NN	81.2
SVC	79.12
MLP	81.6
RL	78.4

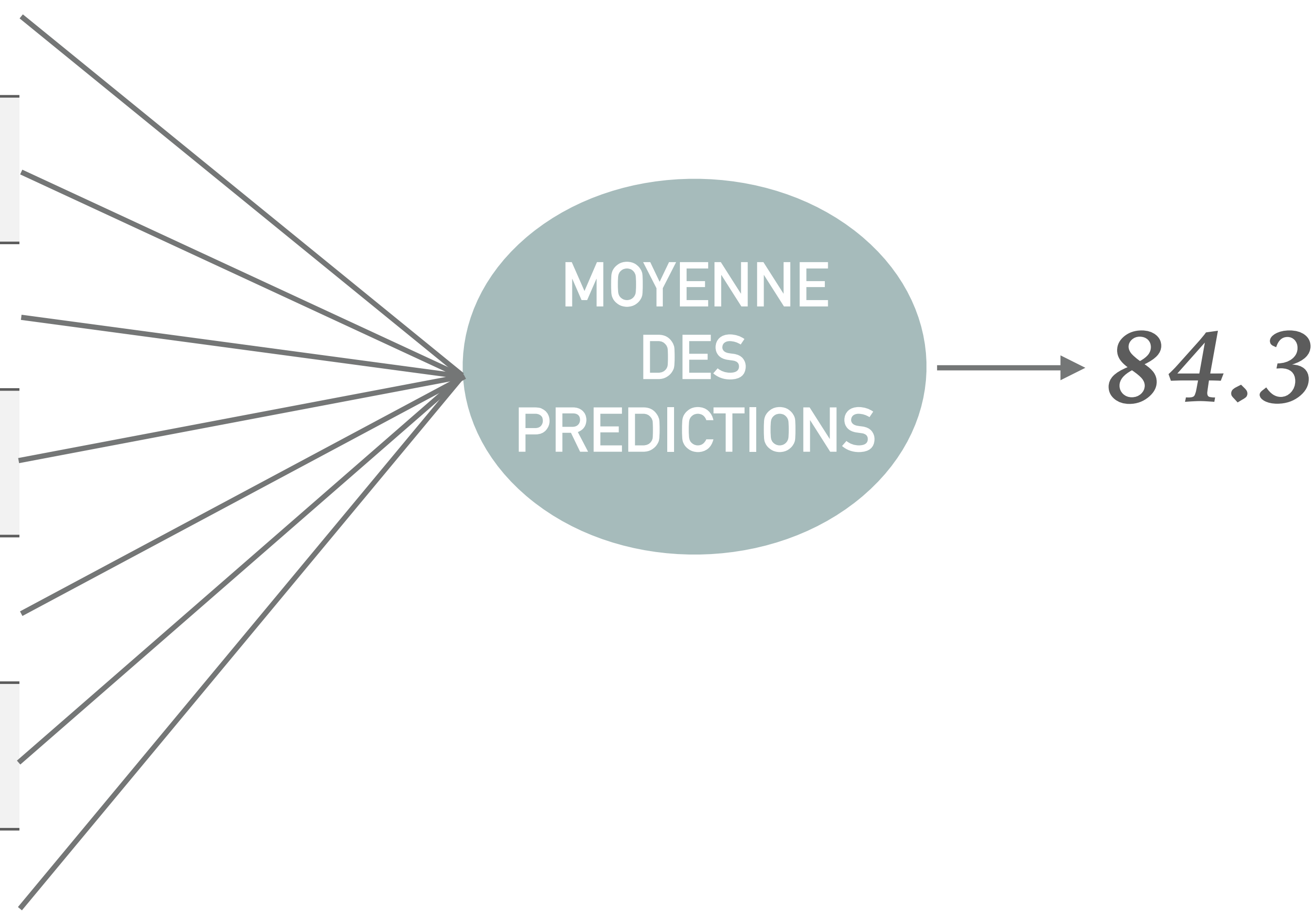
METHODOLOGIE : CREATION DES MODELES

Modèle	Score en VC
RandomForest	82.3
ExtraTrees	82.7
Xgboost	83.6
k-NN	81.2
SVC	79.12
MLP	81.6
RL	78.4



METHODOLOGIE : CREATION DES MODELES

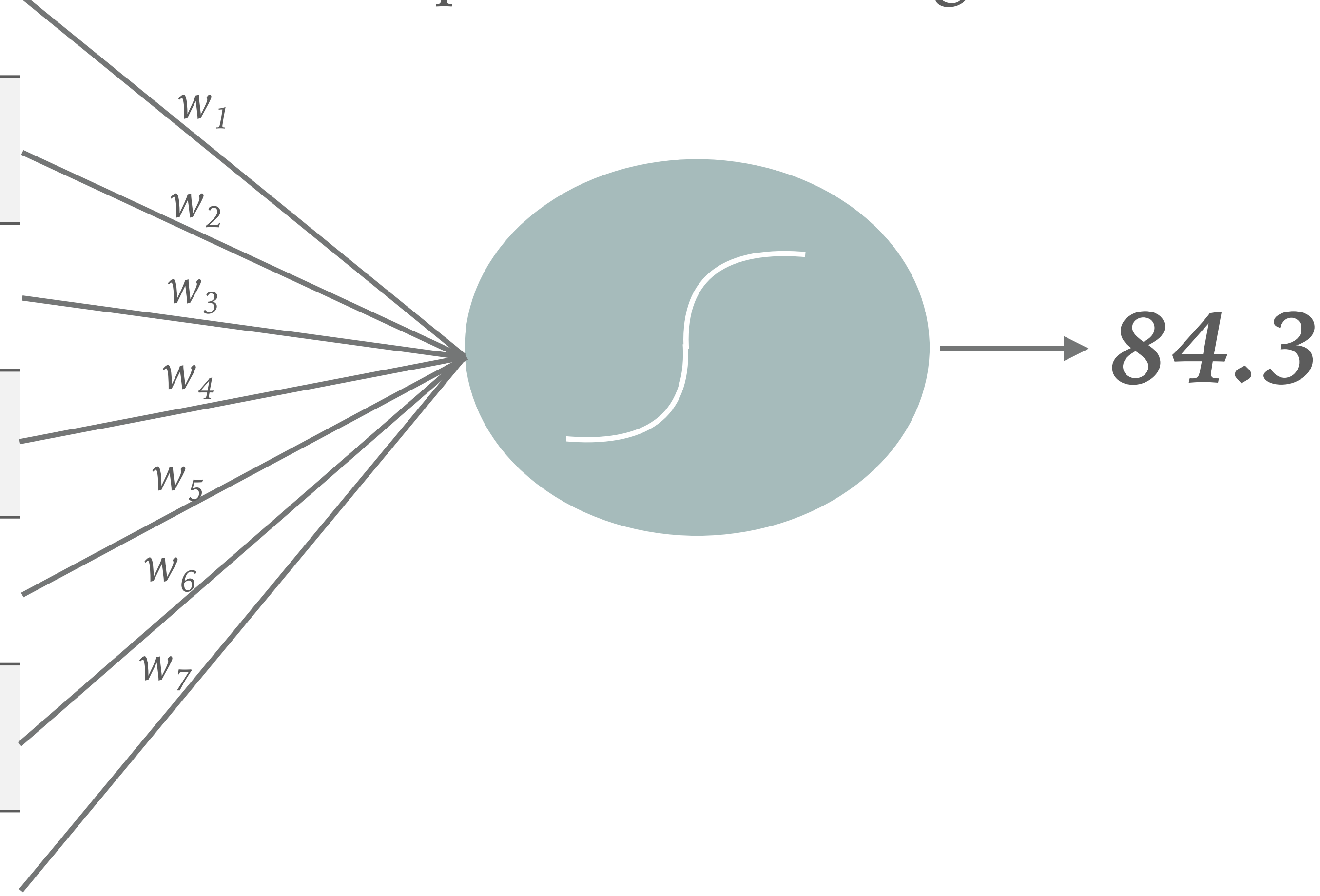
Modèle	Score en VC
RandomForest	82.3
ExtraTrees	82.7
Xgboost	83.6
k-NN	81.2
SVC	79.12
MLP	81.6
RL	78.4



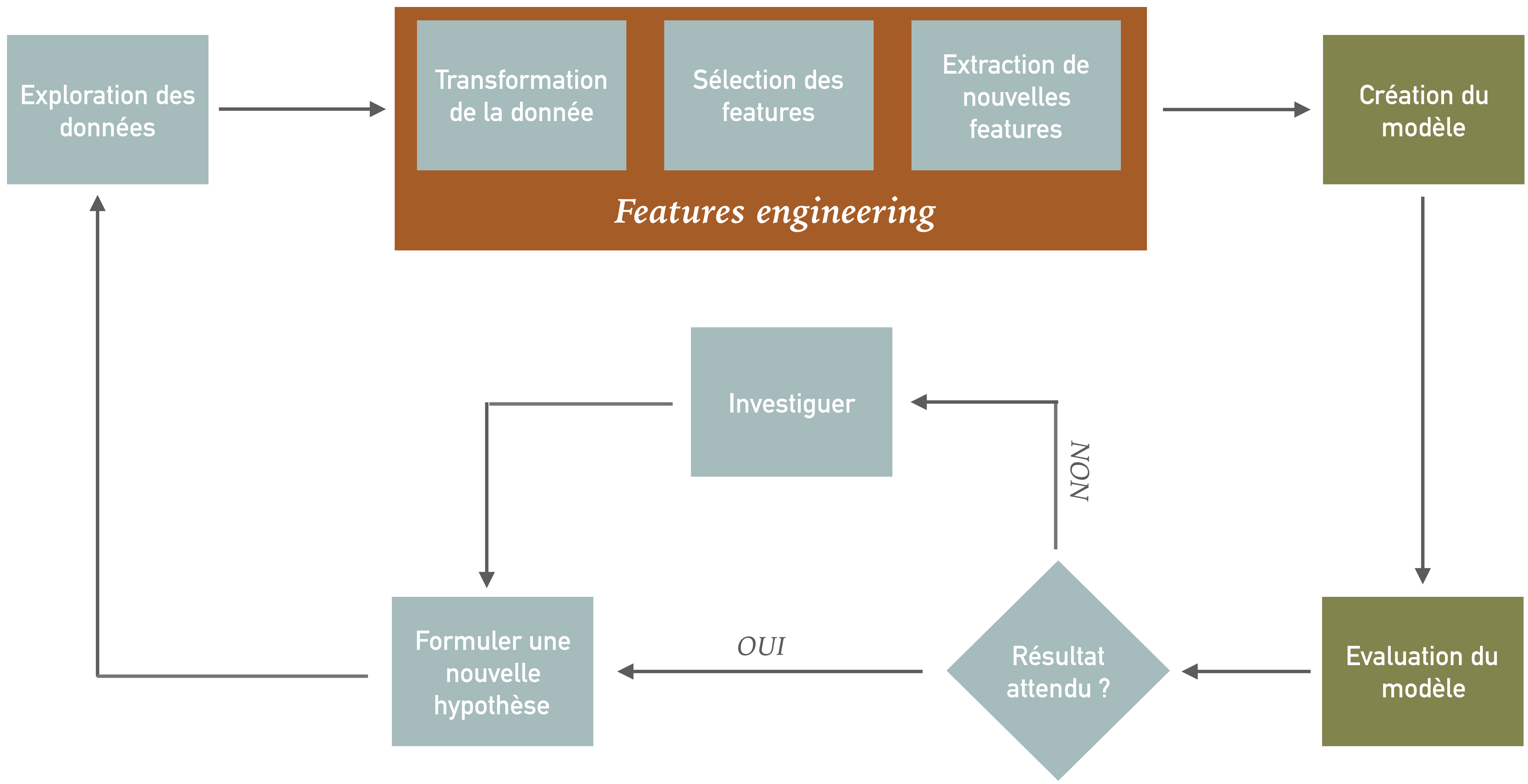
METHODOLOGIE : CREATION DES MODELES

Modèle	Score en VC
RandomForest	82.3
ExtraTrees	82.7
Xgboost	83.6
k-NN	81.2
SVC	79.12
MLP	81.6
RL	78.4

Entraînement d'un nouveau classifieur pour combiner les prédictions des algorithmes



METHODOLOGIE : DEMARCHE EXPERIMENTALE



COMPETITION TITANIC : PREDICTION DES SURVIVANTS

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
465	0	Lower	Maisner, Mr. Simon	male	nan	0	0	A/S 2816	8.05	nan	S
84	0	Upper	Carrau, Mr. Francisco M	male	28	0	0	113059	47.1	nan	S
780	1	Upper	Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)	female	43	0	1	24160	211.338	B3	S
852	0	Lower	Svensson, Mr. Johan	male	74	0	0	347060	7.775	nan	S
748	1	Middle	Sinkkonen, Miss. Anna	female	30	0	0	250648	13	nan	S
362	0	Middle	del Carlo, Mr. Sebastiano	male	29	1	0	SC/PARIS 2167	27.7208	nan	C
200	0	Middle	Yrois, Miss. Henriette ("Mrs Harbeck")	female	24	0	0	248747	13	nan	S
630	0	Lower	O'Connell, Mr. Patrick D	male	nan	0	0	334912	7.7333	nan	Q
732	0	Lower	Hassan, Mr. Houssein G N	male	11	0	0	2699	18.7875	nan	C
175	0	Upper	Smith, Mr. James Clinch	male	56	0	0	17764	30.6958	A7	C
538	1	Upper	LeRoy, Miss. Bertha	female	30	0	0	PC 17761	106.425	nan	C
384	1	Upper	Holverson, Mrs. Alexander Oskar (Mary Aline Towner)	female	35	1	0	113789	52	nan	S
881	1	Middle	Shelley, Mrs. William (Imanita Parrish Hall)	female	25	0	1	230433	26	nan	S
173	1	Lower	Johnson, Miss. Eleanor Ileen	female	1	1	1	347742	11.1333	nan	S
327	0	Lower	Nysveen, Mr. Johan Hansen	male	61	0	0	345364	6.2375	nan	S

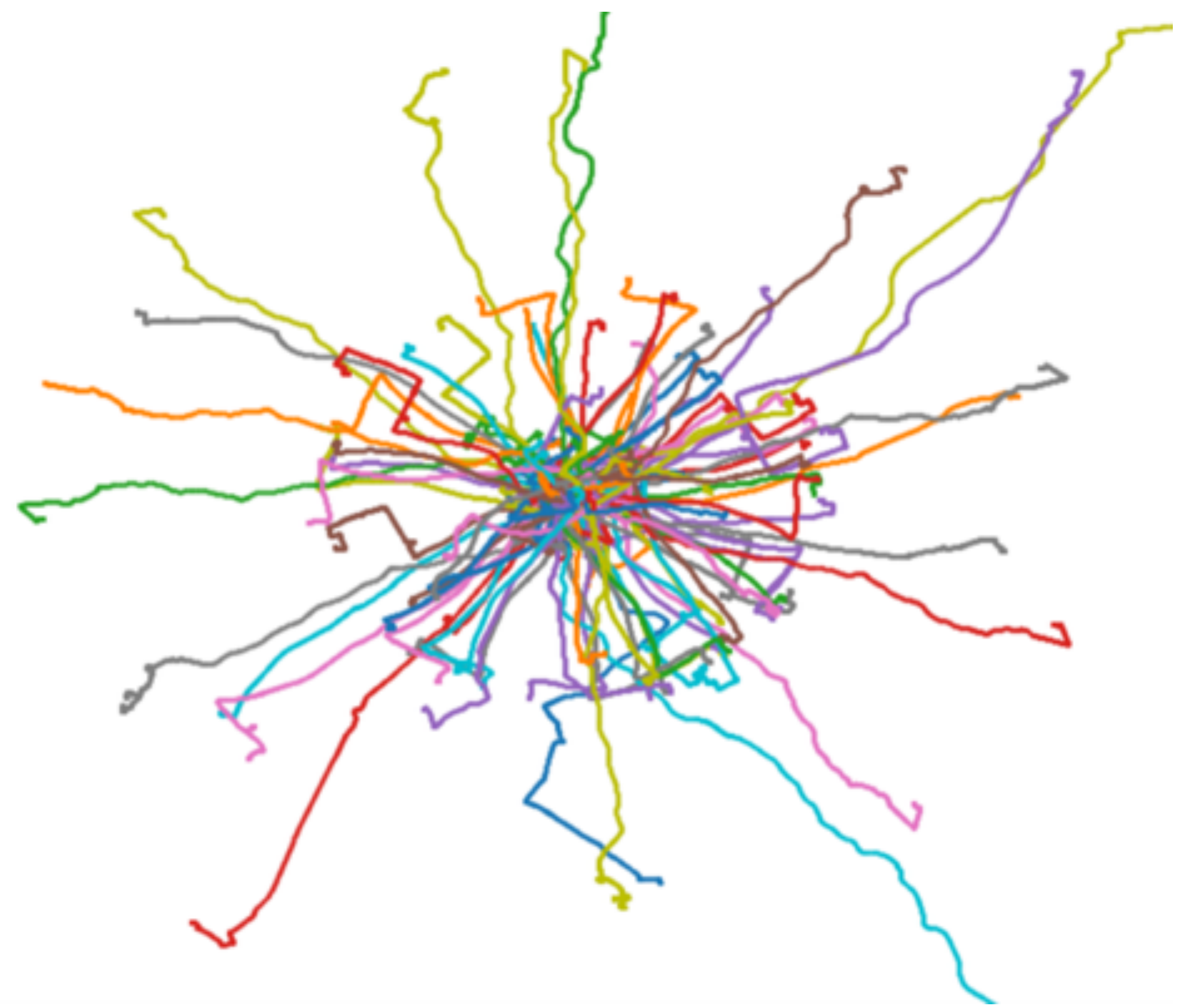
COMPETITION AXA : PREDICTION DU CONDUCTEUR

	x	y
0	0	0
1	-1.9	
2	-3.5	
3	-5.4	
4	-7.4	
5	-9.1	
6	-10.8	
7	-11.9	
8	-12.8	
9	-13.2	
10	-12.4	
11	-9.8	
336	497.3	
337	497.3	
338	497.3	
339	497.3	
340	497.3	

	x	y
0	0	0
1	0	
2	-0.1	
3	-0.1	
4	-0.4	
5	-1.1	
6	-2.2	
7	-3.3	
8	-4.1	
9	-4.5	
10	-4.9	
11	-4.9	
798	9918.4	
799	9918.2	
800	9918	
801	9918	
802	9918.1	

	x	y
0	0	0
1	-2.3	
2	-3.9	
3	-5.8	
4	-7.2	
5	-5.9	
6	-3.9	
7	-1.6	
8	1.8	
9	5.5	
10	9.7	
11	13.9	
940	-3544	[illegible]
941	-3544	[illegible]
942	-3544	[illegible]
943	-3544	[illegible]
944	-3544	[illegible]

	x	y
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0.1	-0.1
7	0.1	-0.1
8	0.1	-0.1
9	0.1	-0.1
10	-0.5	-1.1
11	-0.4	-1.1
487	908.9	1407.5
488	908.8	1407.5
489	908.8	1407.5
490	908.8	1407.5
491	909	1407.4



QUESTIONS

QUESTIONS ???