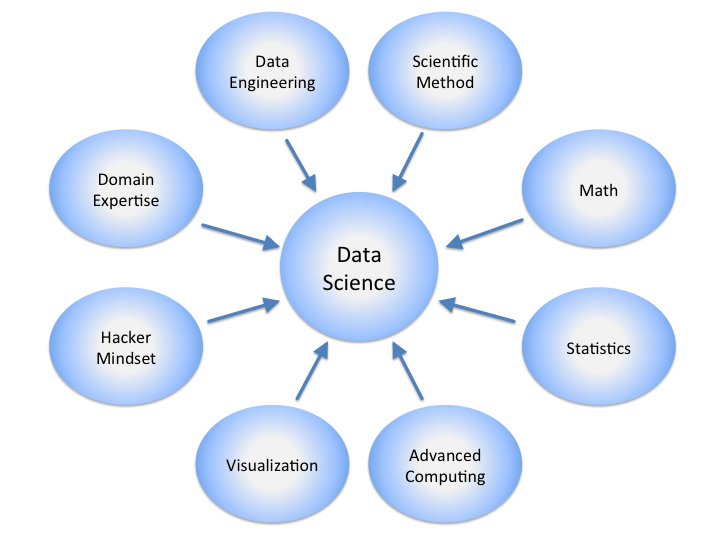
Développer une preuve de concept

## I. Introduction

#### A. Introduction au domaine du machine learning et de la data science qui évolue rapidement



La data science ou “Science des données” est une science qui regoupe plusieurs disciplines : les mathématiques, les statistiques, la visualisation, le machine learning etc

Le machine learning ou “apprentissage machine” en francais, est la discipline qui va permettre a l’ordinateur d’apprendre a partir d’un jeu de données. On pourra l’utiliser pour faire des prédictions, établir des tendances ou même prendre des décisions en fonction des données.

La data science est en plein essor, de plus en plus d'entreprises font appel à elle. Elle est utilisée dans plein d'industries, de la finance à la santé. Elle aide les gouvernements à détecter les fraudes fiscales, à gérer les risques dans les fonds d'investissement, à trouver les meilleures chances de réussite pour soigner les malades à l'hôpital, etc.

On peut la mettre en place dans presque tous les secteurs, même le BTP par exemple pour prédire la consommation d'énergie à l'avance.

#### B. Objectif du projet

L’objectif du projet est d’améliorer la méthode actuallement utilisée en production chez un client pour prédire le taux de survie des passagers du titanic. Pour ce faire, nous allons explorer de nouvelles avancées.

Le choix du nouvel algorithme a tester c’est porté sur le quanticSVM.

Le SVM (support vector machine) est une technique de classification qui est largement utilisée.

Dans ce travail, on s’interesse a une version quantique du SVM.

La litterature scientifique a demontre que dans certaines circonstance les algorithmes quantiques permettent d’ameliorer la performance de la classification.

Cependant, dans la plupart cas, les evaluations ont ete realise avec des données artificielles (pas reelles, créées dans un but d’experimentation). Par exemple:

Liu, Y.; Arunachalam, S.; Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **2021**, *17*, 1013–1017.

Nous allons essayer d’appliquer ces methodes a un vrai dataset (donc celui du Titanic)...

nous allons comparer la performance du SVR avec un noyau quantique et un noyau RBF. On comparera egalement la performance des SVR avec une regression lineaire simple.

On s’inspire du travail realisé dans cette article, et qui presente une classification quantique de VRAIES données (dans cette article il s’agit de données encepahlographiques).

[Brain Sciences | Free Full-Text | Case-Based and Quantum Classification for ERP-Based Brain&ndash;Computer Interfaces (mdpi.com)](https://www.mdpi.com/2076-3425/13/2/303)

#### C. Notes techniques (quelle librairie utiliser?, etc..)

Dans ce travail on utilise la librairie pyRiemann-qiskit qui est une sur-couche autour de la librairie Qiskit.

Qiskit est une bibliothèque quantique de pyRiemann.

Elle a été conçue pour être aussi proche que possible des interfaces de sklearn. Elle a pour objectif de réduire l'écart entre les algorithmes classiques et les algorithmes quantiques.

L’interet de la surcouche pyRiemann-qiskit, c’est de permettre de configurer plus facilement la classifieur.

Elle permettrait aussi de classifier, non seulement des vecteurs, mais aussi des matrices, grace a une methode appellée la “geoemtrie Riemannienne”. Pour les lecteurs interessés, on conseille l’article: …

M. Congedo, A. Barachant, and R. Bhatia, ‘Fixed Point Algorithms for Estimating Power Means of Positive Definite Matrices’, *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2211–2220, avril 2017, doi: [10.1109/TSP.2017.2649483](https://doi.org/10.1109/TSP.2017.2649483).

## II. Choix des données

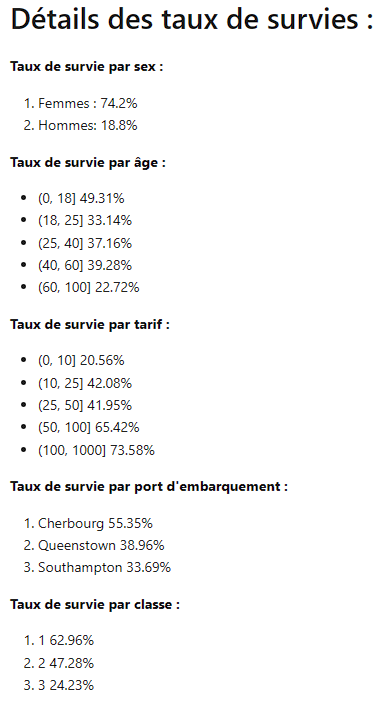
#### 1. Explication du choix du projet utilisé comme référence

Nous avons choisi avec mon mentor de reprendre notre projet Titanic dans l’objectif de potentiellement améliorer notre score SVC avec noyau RBF avec le QuanticSVM de pyRiemann.

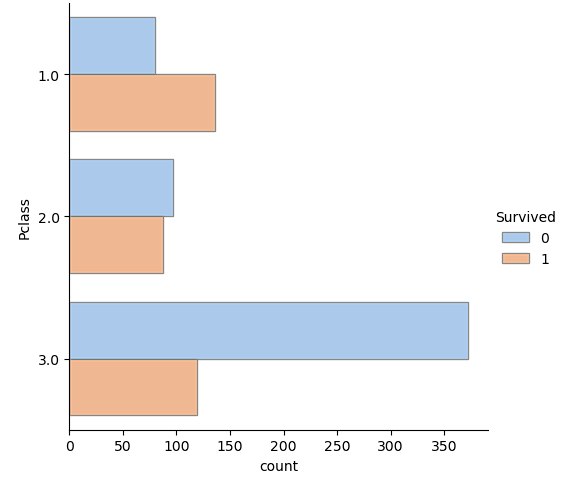
#### 2. Description du jeu de données sélectionnées

Le projet Titanic est a la base une compétition Kaggle, dont l’objectif est de prédire le taux de survie des passagers du nauffrage. Pour ce faire, on utilise plusieurs variables.

Nous avons calculer l’accuracy avec Regression linéaire et SVC puis le score SVC avec noyau RBF.

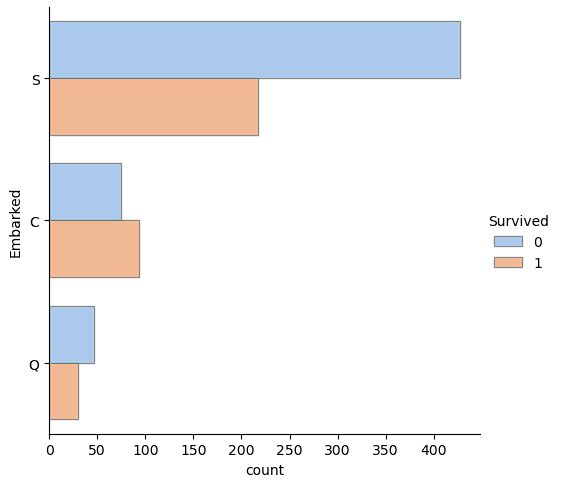


**Details des taux de survie par Classe**



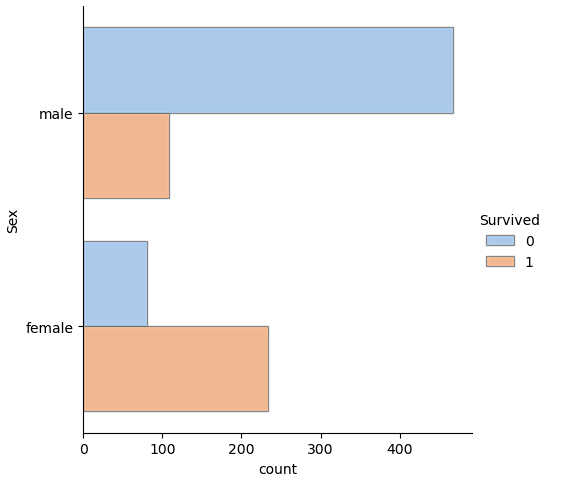
On notera que le taux de mortalité est beaucoup plus élevé par la classe 3. Qu’il y a presque autant de décés que de survivants dans la classe 2 et un taux de survie beaucoup plus élevé pour la classe 1.

**Taux de survie par port d’embarquement**



Le port d’embarquement de Southampton a eu le taux de décés le plus élevé. Le port d’embarquement de Cherboug a eu le taux de décés le plus bas. le port de Queenstown a eu un taux un taux de décés élevé par rapport aux nombre de passagers qui y ont embarqué.

**Taux de survie par genre :**



On constate que le taux de décés est significativement plus élevé pour les hommes que pour les femmes.

## III. Choix de l'algorithme implémenté en prototype

#### A. Recherche d'un article de recherche récent

**[1]**

**A. Andreev, G. Cattan, S. Chevallier, and Q. Barthélemy, ‘pyRiemann-qiskit: A Sandbox for Quantum Classification Experiments with Riemannian Geometry’, *Research Ideas and Outcomes*, vol. 9, Mar. 2023, doi:** [**10.3897/rio.9.e101006**](https://doi.org/10.3897/rio.9.e101006)**.**

**[1]**

**V. Havlíček *et al.*, ‘Supervised learning with quantum-enhanced feature spaces’, *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019, doi:** [**10.1038/s41586-019-0980-2**](https://doi.org/10.1038/s41586-019-0980-2)**.**

#### B. Description de l'algorithme sélectionné et de ses techniques

L'informatique quantique est une nouvelle technologie prometteuse pour le Machine Learning qui peut être moins couteuse et potentiellement donner de meilleurs résultats.

Nous avons décider d’utiliser Qiskit, une bibliothèque d’’informatique quantique développée par IBM, et pyRiemann, qui est un cadre qui aide à analyser les signaux du cerveau. La bibliotheque pyRiemann se base sur la théorie des espaces de Riemann pour extraire des caractéristiques discriminantes à partir de ces signaux et les utiliser dans des tâches de classification.

La fusion de ces bibliothéques a permis la création d'un processus standardisé appelé "QuantumClassifierWithDefaultRiemannianPipeline" qui permet de classer les ondes cérébrales en deux catégories.

Cette fusion combine les avantages de l'informatique quantique avec l'analyse des signaux cérébraux pour créer un modèle de classification robuste et performant.

Comment l’algorythme fonctionne t’il ?

La bibliotheque pyRiemann va analyser les signaux de cervaux tels que les électroencéphalogrammes (EEG) ou les électromyogrammes (EMG). L'EEG mesure l'activité électrique du cerveau, tandis que l'EMG mesure l'activité électrique des muscles.

Il va donc y avoir plusieurs étapes : la normalisation des données, la sélection des caractéristiques pertinentes et la réduction de dimension.

Puis Qiskit est utilisé pour construire le modèle de classification. Afin de classifier les signaux cérébraux, Qiskit créé des instructions pour manipuler les "qubits", qui sont les unités de base de l'ordinateur quantique. Les qubits sont utilisés pour stocker les caractéristiques des signaux cérébraux et effectuer les calculs nécessaires pour la classification. Ils permettent de réaliser des opérations complexes plus rapidement qu’un ordinateur normal.

Après avoir construit le modèle de classification quantique, nous utilisons les données des signaux cérébraux pour entraîner le modèle. Le modele peut ensuite être exécuté sur un ordinateur quantique ou sur un simulateur Qiskit.

#### C. Justification du choix de l'algorithme en fonction de ses avantages et de ses résultats

Il semblerait que l'informatique quantique offre des avantages par rapport à l'informatique classique, notamment en termes de vitesse de calculs et de résultats.

Cependant, l'efficacité dépend des données utilisées, de leur qualité et de la manière dont elles sont encodées en états quantiques.

Le QuanticSVM est tout simplement un SVC avec noyau quantic.

## IV. Les méthodes utilisées dans le projet Titanic

#### A. La régréssion linéaire

Nous avons initialement utilisé la Regression linéaire.

La régression linéaire est une technique d'analyse de données qui va nous permettre de prédire la valeur de données inconnues en utilisant une autre valeur de données que l’on connait déja.

La regression linéaire va établir une relation entre deux variables, une connue et une inconnue en tracant un graphique avec une variable indépendante (x) et une variable dépendante (y). L’apprentissage automatique va chercher à trouver une ligne droite qui représente au mieux la relation entre ces deux variables. Il va ensuite chercher la droite qui represente le mieux la relation entre ces variables.

Elle est trés utile pour établir des prévisions, et est donc trés adaptée aux prévisions de taux de survie des passagers du Titanic.

#### B. Le SVR (machines à vecteurs de support) avec noyau RBF ( Radial Basis Function)

Nous avons également utilisé le SVR, ou Support Vector Regression, qui est une méthode de machine learning, qui comme la régréssion linéaire va permettre de prédire des valeurs inconnues a l’aide de valeurs connues.

Cependant au lieu de tracer une ligne droite comme la régression linéaire, le SVR fonctionne en utilisant des vecteurs de support. La technique "noyau" permet de transformer les données pour les rendre linéairement séparables et ainsi de modéliser des relations non linéaires entre les variables.

Pour faire court, le SVR peut prendre en compte des relations entre x et y qui ne peuvent pas être capturées par une simple ligne droite, c’est un modele qui est plus adapté lorsque les relations entre les variables sont plus compliquées.

#### C. Methode

D’abord on transforme nos données dans un etat quantique avec une ZZFeatureMap.

Pour ce travail, on a mis le nombre de repetition a 2 et on a utilisé 1024 “shots”. Un shot c’est tout simplement le nombre de fois ou est executé le circuit quantique pour la classification.

En effet, dans un ordinateur quantique le fait de mesurer un resultat resulte en une perte d’information. Le resultat est donc en parti “errone”. Afin de supprimer ce biais, une maniere de proceder est d’execulté l’algorithme plusieurs fois.

Nous n’avons pas de vrai ordinateur quantique mais nous avons emule l’ordinateur avec QasmSimulator. QasmSimulator est a l’ordinateur quantique ce que l’assembleur est pour l’ ordinateur classique.

#### D. Résultats

|  | **Résultat** |
| --- | --- |
| **Régréssion Linéaire** | 0.8044 |
| **SVC** | 0.5977 |
| **SVC avec noyau RBF** | 0.5175 |
| **QuanticSVM** | **0.7707** |

Les étapes :

1. Nous avons donc commencé par la Régréssion logistique qui nous a donné un score d’accuracy de 0.8044.
2. Puis nous avons mis en place un SVC de scikit learn avec lequel nous avons obtenu un score de 0.5977.
3. Aprés le SVC nous avons testé avec un modele SVC avec noyau RBF, dans lequel nous avons obtenu un score de 0.5175. Le score est inférieur au SVC, on peut donc penser que les modeles linéaires marchent mieux parce que le probleme est linéaire.
4. Et enfin nous avons testé a l’aide du QuanticSVM de la librairie pyriemann\_qiskit, cependant le temps de chargement du modele est beaucoup plus long que le SVC. Aprés deux jours de chargement, nous avons obtenu un score de 0.7707.

#### 

Le modele QuanticSVM nous donne donc un score supérieur au SVC de scikit learn, ou même du modele SVC avec noyau RBF

## V. Conclusion

#### A. Perspectives et attentes pour le projet, en mettant en évidence les défis et les opportunités

Le QuanticSVM est donc une alternative au SVC, au SVC avec noyau RBF et a la régréssion linéaire. Cependant on est limité par le temps, en effet le QuanticSVM est plus adapté aux ordinateurs quantiques. Le temps de chargement est beaucoup trop long par rapport aux autres modeles sur un PC lambda.

A l’heure actuelle seule une poignée d’entreprise possede des ordinateur quantiques ( Google, IBM, Microsoft…). Pourtant certaines entreprises comme IBM avec IBM Quantum Experience proposent l’accés a ses ordinateurs quantiques via le cloud, et donc accéssible a n’importe quelle entreprise ou même a un usage individuel.

Avec l’informatique Quantique on pourra donc réduire le temps de calculs tout en améliorant nos résultats.

## VI. Sources bibliographiques

#### A. Références bibliographiques pertinentes

[Sklearn wrapper around QSVM and VQC · Issue #240 · qiskit-community/qiskit-machine-learning · GitHub](https://github.com/qiskit-community/qiskit-machine-learning/issues/240)

[GitHub - pyRiemann/pyRiemann-qiskit: A library for machine learning and quantum programming based on pyRiemann and Qiskit projects](https://github.com/pyRiemann/pyRiemann-qiskit)

[pyRiemann-qiskit/classification.py at main · pyRiemann/pyRiemann-qiskit · GitHub](https://github.com/pyRiemann/pyRiemann-qiskit/blob/main/pyriemann_qiskit/classification.py)

[pyriemann\_qiskit.classification.QuanticSVM — pyRiemann-qiskit 0.0.3 documentation](https://pyriemann-qiskit.readthedocs.io/en/latest/generated/pyriemann_qiskit.classification.QuanticSVM.html#pyriemann_qiskit.classification.QuanticSVM)

[Brain Sciences | Free Full-Text | Case-Based and Quantum Classification for ERP-Based Brain&ndash;Computer Interfaces (mdpi.com)](https://www.mdpi.com/2076-3425/13/2/303)