



Business Intelligence and User Modeling

Rapport de projet

Le suicide ! quel est le lien entre la situation socio-économique d'un pays et son taux de suicide ?

Étudiantes:

Sarah DJOUDER
Tinhinane CHAFAI

Numéros:

3970876
3802927

Encadré par:

Nicoals Castanet
Laure Soulier

Mai 2022

Table des matières

1	Introduction	1
1.1	Problématiques	1
1.2	Sources de données :	1
2	Modélisation: Schéma conceptuel	2
3	Exploration et analyses de données	3
3.1	Analyse des suicides par pays (Dimension Lieu)	3
3.2	Analyse des suicides par sexe et âge	4
3.3	Analyse des suicides par années (Dimension Temporelle)	4
3.4	Analyse des suicides par dimension socio-économique :	5
4	Modèle Machine Learning	7
4.1	L'importance des features par modèles de classification :	8
5	Synthèses	10
5.1	Les facteurs socio-économique favorisant les suicides :	10
5.2	Prévention	10

1 Introduction

Chaque année, près de 703 000 personnes se suicident et beaucoup d'autres font une tentative de suicide. Chaque suicide est une tragédie qui touche les familles, les communautés et des pays entiers et qui a des effets durables sur ceux qui restent. Le suicide intervient à n'importe quel moment de la vie et était deuxième cause de mortalité chez les jeunes dans le monde. Le suicide est un grave problème de santé publique ; or il peut être évité moyennant des interventions menées en temps opportun, fondées sur des données factuelles et souvent peu coûteuses. Pour que l'action nationale soit efficace, une stratégie globale multisectorielle de prévention du suicide s'impose.

1.1 Problématiques

Le suicide est évitable. Pourtant, toutes les 40 secondes, une personne se suicide quelque part dans le monde et bien plus tentent de mettre fin à leurs jours. Aucune région ni aucune tranche d'âge n'est épargnée. Il touche toutefois particulièrement les jeunes de 15 à 29 ans, chez qui il constitue la deuxième cause de mortalité à l'échelle mondiale.

Qu'est-ce qui pousse une personne à se suicider ? Pourquoi les gens sont-ils chaque année si nombreux à mettre fin à leurs jours ? Quels sont les facteurs en cause ? La pauvreté ? Le chômage ? La rupture amoureuse ? Ou cela peut-il s'expliquer par une dépression ou d'autres troubles mentaux ? Le suicide est-il un acte impulsif ou est-il favorisé par les effets désinhibants de l'alcool ou de la drogue ? Les questions se bousculent, mais il n'est pas simple d'y répondre sommairement.

Le comportement suicidaire est un phénomène complexe influencé par plusieurs facteurs interdépendants (personnels, sociaux, psychologiques, culturels, biologiques et environnementaux). À ce jour, seuls quelques pays ont inscrit la prévention du suicide au nombre de leurs priorités sanitaires et 38 pays seulement déclarent s'être dotés d'une stratégie nationale de prévention du suicide. Il est important pour les pays de mieux sensibiliser la communauté et de faire tomber ce tabou afin de faire progresser la prévention du suicide.

Dans le cadre de notre projet on va essayer de répondre à l'une de ces problématiques : **quel est le lien entre la situation socio-économique d'un pays et son taux de suicide ?**

1.2 Sources de données :

Au niveau mondial, on manque de données sur le suicide et les tentatives de suicide, et celles-ci ne sont pas de bonne qualité. Seuls 80 États Membres environ disposent de systèmes d'enregistrement des données d'état civil de qualité qui puissent être utilisés directement pour estimer les taux de suicides. Mais compte tenu de la sensibilité de cette question – et de l'illégalité du comportement suicidaire dans certains pays :

1. **World health organization** : Des datasets contenant des données sur le taux de suicide sur chaque pays dans le monde (le taux de suicide et la moyenne de suicide par 100k de population au cours de quelques années.)
2. **Data World**: Des données économiques sur les pays du monde (Sigle de produit intérieur brut de chaque pays, rémunération des employés,...)
3. **Épidémiologie dans le monde** : Les suicides dans le monde (le taux d'évolution,...)
4. **Web scraping** : On a récupéré les données liées à la religion de chaque pays (pourcentage d'importance de la religion dans un pays) sur des tableau wikipedia

2 Modélisation: Schéma conceptuel

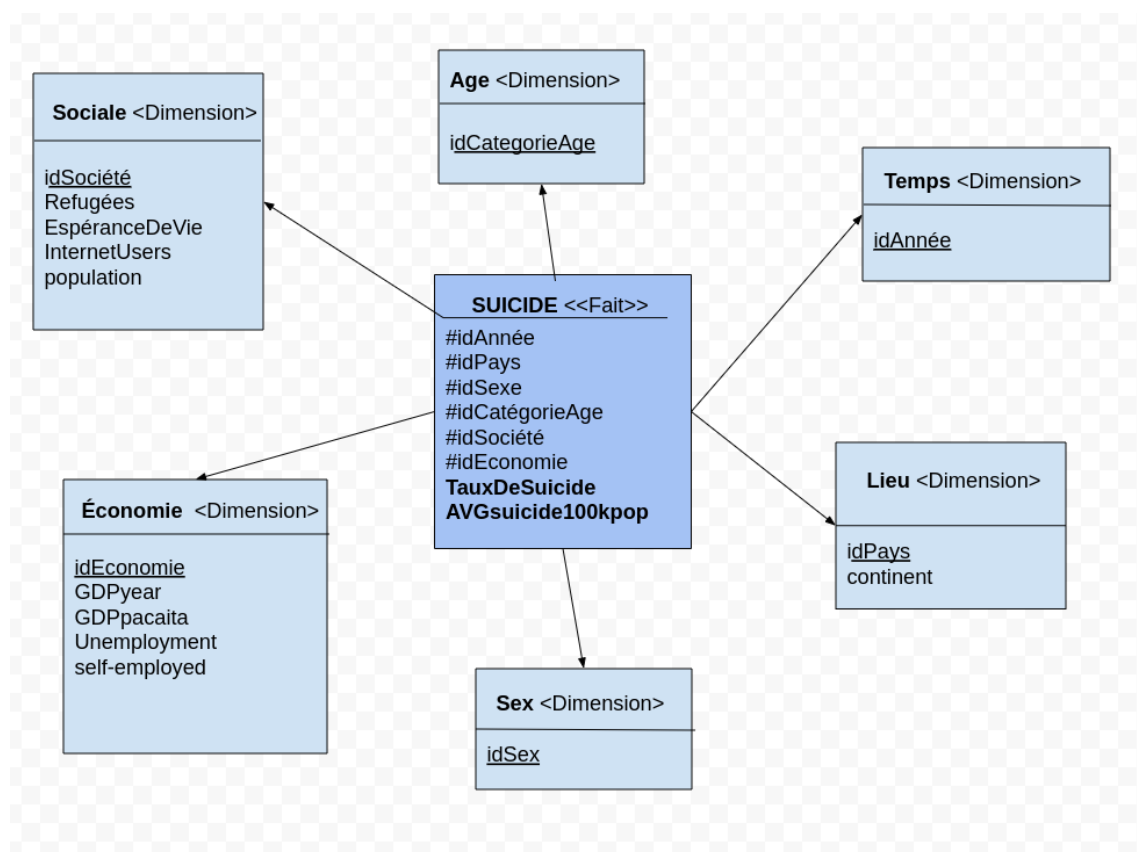


Schéma en étoile caractérisant le fait et les dimensions

3 Exploration et analyses de données

3.1 Analyse des suicides par pays (Dimension Lieu)

Les pays avec le plus de suicides

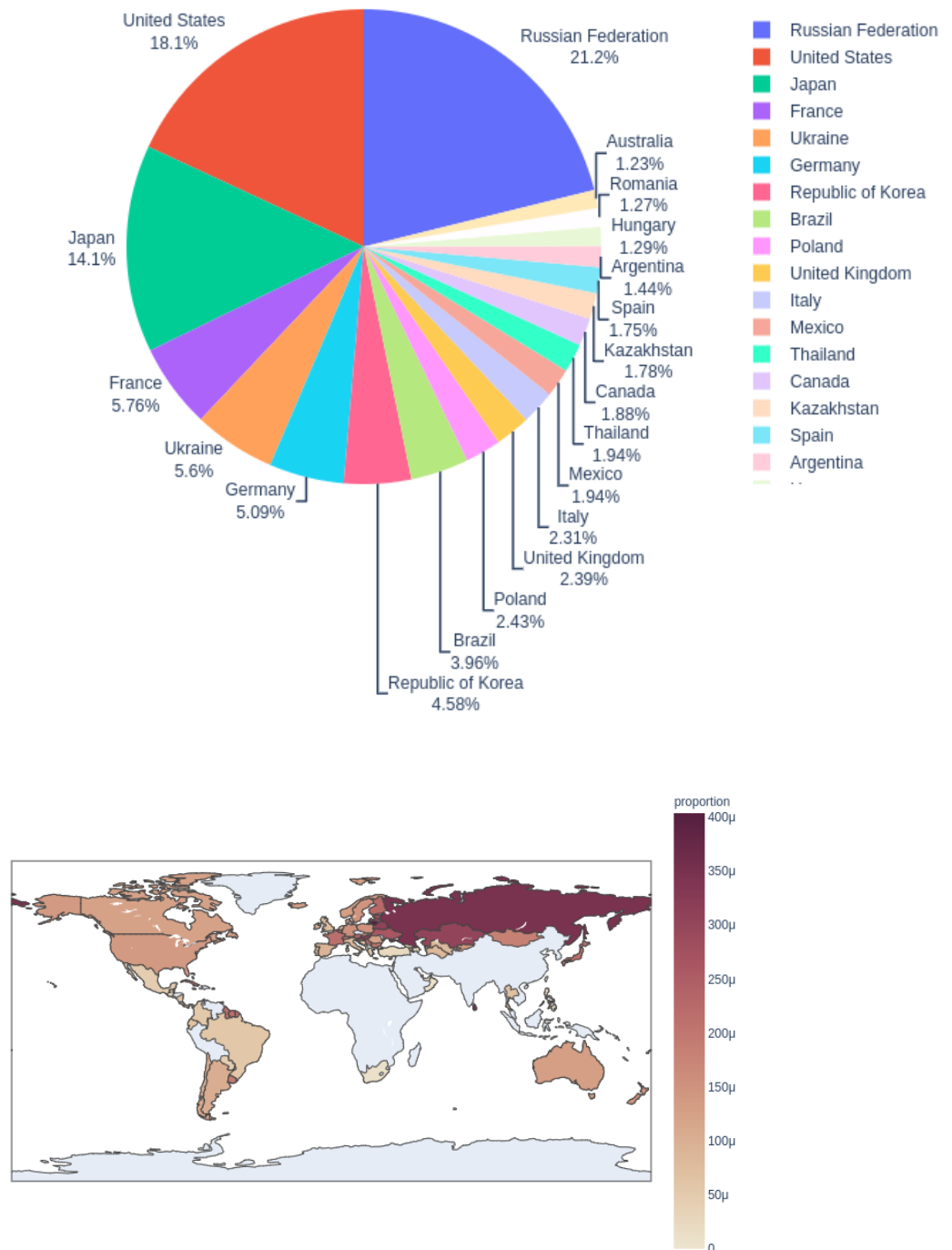


Figure 3.1: Distribution globale des suicides dans le monde

L'analyse de notre dataset choisi montre que la fédération russe, les état-unis ainsi que le japon sont les trois pays ayant le plus taux de suicide avec 21,2%, 81,1% et 14,1% respectivement.

3.2 Analyse des suicides par sexe et âge

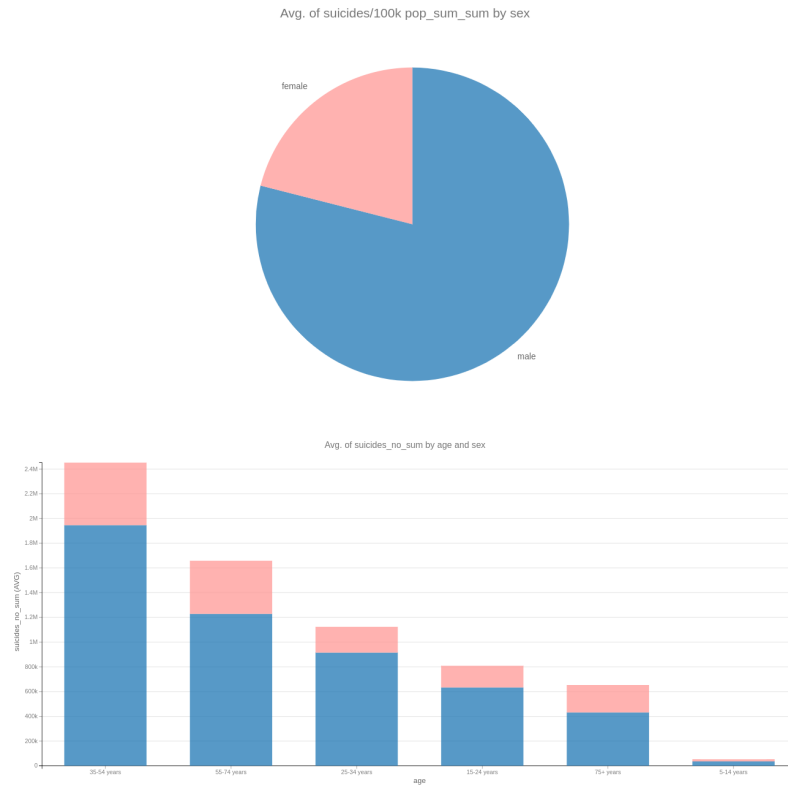


Figure 3.2: Taux de suicides par sexe et par âge

La figure 3.2 représente le taux de suicides par sexe et âge. On remarque que le taux de suicide de la catégorie hommes est beaucoup plus élevé par rapport à celui des femmes, spécifiquement pour les personnes entre 35 et 54 ans.

3.3 Analyse des suicides par années (Dimension Temporelle)

La figure 3.4 représente les proportions de suicide à travers les années. L'année 1995 a marqué le taux le plus élevé avec une proportion de 250μ pour les hommes et plus de 50μ pour les femmes. De même, une analyse a été effectuée sur la proportion de suicide en 100k à travers les années pour chaque tranche d'âge comme le montre la figure 3.3. L'étude a montré que les personnes ayant plus de 75 ans ont le taux de suicide le plus élevé.

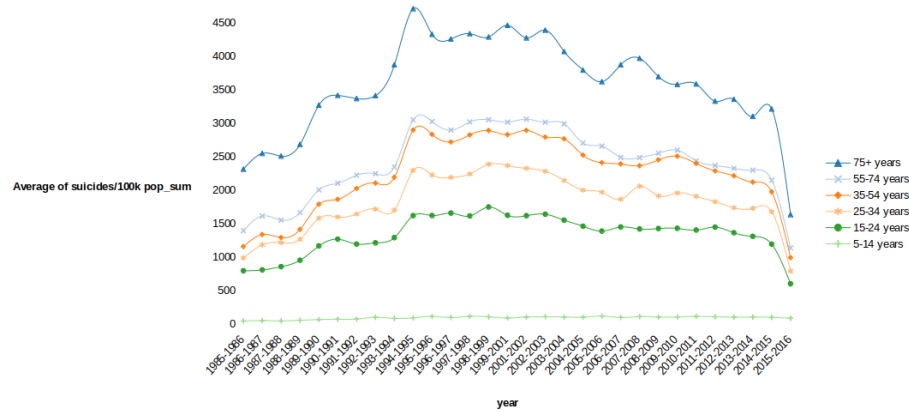


Figure 3.3: Analyse des suicides par année et catégorie d'âge

Homme v/s Femme suicide proportion a travers les années

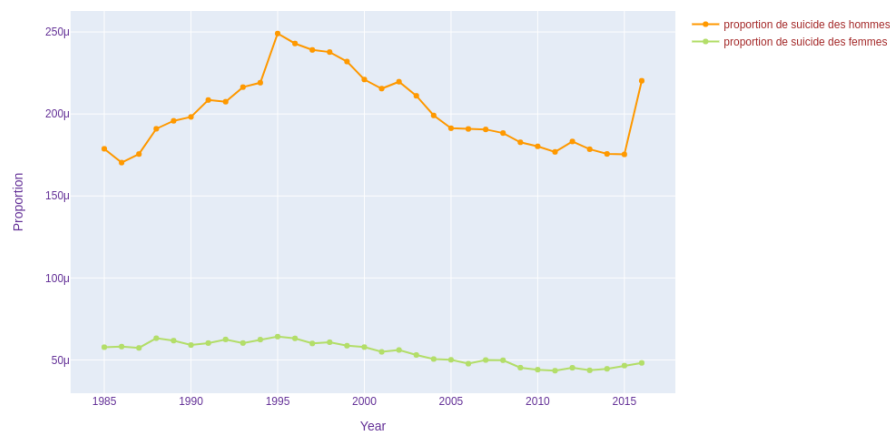


Figure 3.4: Analyse des suicides par années

3.4 Analyse des suicides par dimension socio-économique :

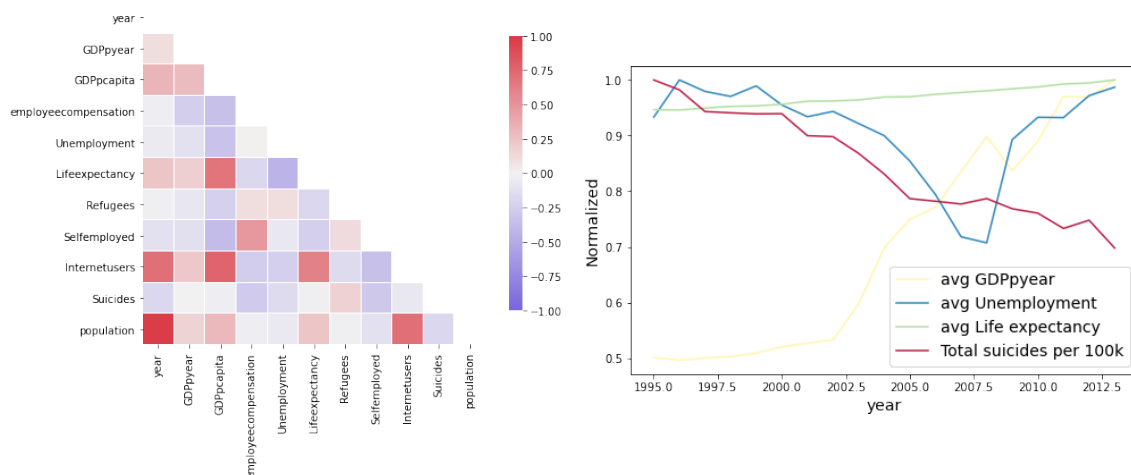


Figure 3.5: Corrélation entre le suicide et les variables socio-économiques citées.

On a essayé de visualisé les corrélations entre le taux de suicide et les différentes autres valeurs socio-économique. Les analyses qu'on pourra tirer de ces deux figures:

- Le taux de suicide représenté en rouge décends au fil des années.
- Le GDP représenté en jaune est entrain de monter au fil des années.
- On voit bien la corrélation inverse entre ces deux mesure: plus le GDPpyear est entrain de monter, le taux de suicide est entrain de diminuer.
- Le taux de chômage decends avec la même pente que le taux de suicide au cours des 15 premières années (la courbe en bleue et la courbe en rouge sur le graphe).

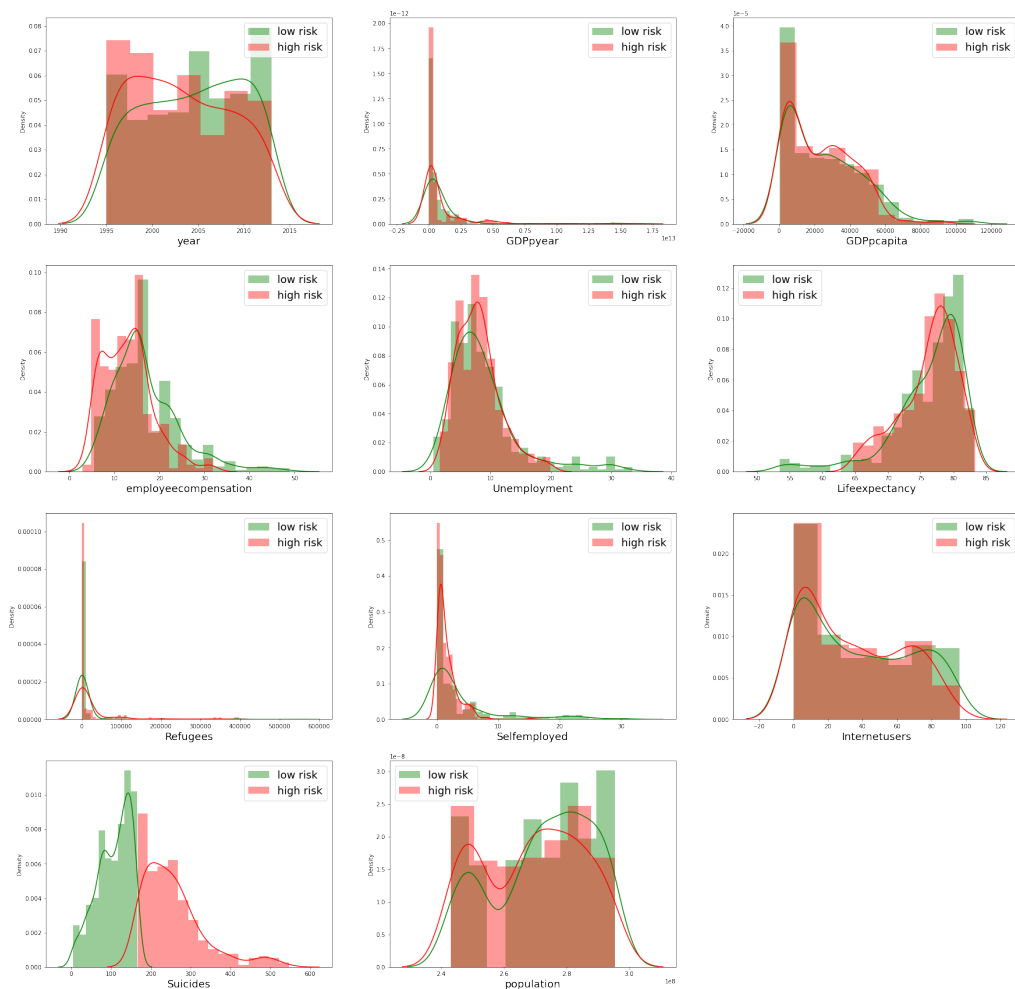


Figure 3.6: Distribution des variables socio-économiques dans les 2 mondes : grand taux de suicide et bas taux de suicide

4 Modèle Machine Learning

Dans le but de répondre à notre problématique on a procédé par une classification binaire :

1. **Classe 1** : Tous les pays avec un taux de suicide **plus grand que la moyenne**.
2. **Classe 2** : Tous les pays avec un taux de suicide **plus petit que la moyenne**.

Tourner différents modèles de Machine Learning pour séparer au mieux les deux classes et regarder en fin les features avec le plus **grand poids contribuant à cette séparation**.

Pour cela on a fait une petite campagne d'expériences avec les modèles suivant :

1. Régression Logistique
2. Arbres de Décision
3. Random Forests
4. Perceptron Multi-Couches

On a regardé les performances de chacun des modèles en Cross validation. Les figures suivantes montrent les courbes Roc obtenues pour chacun des modèles :

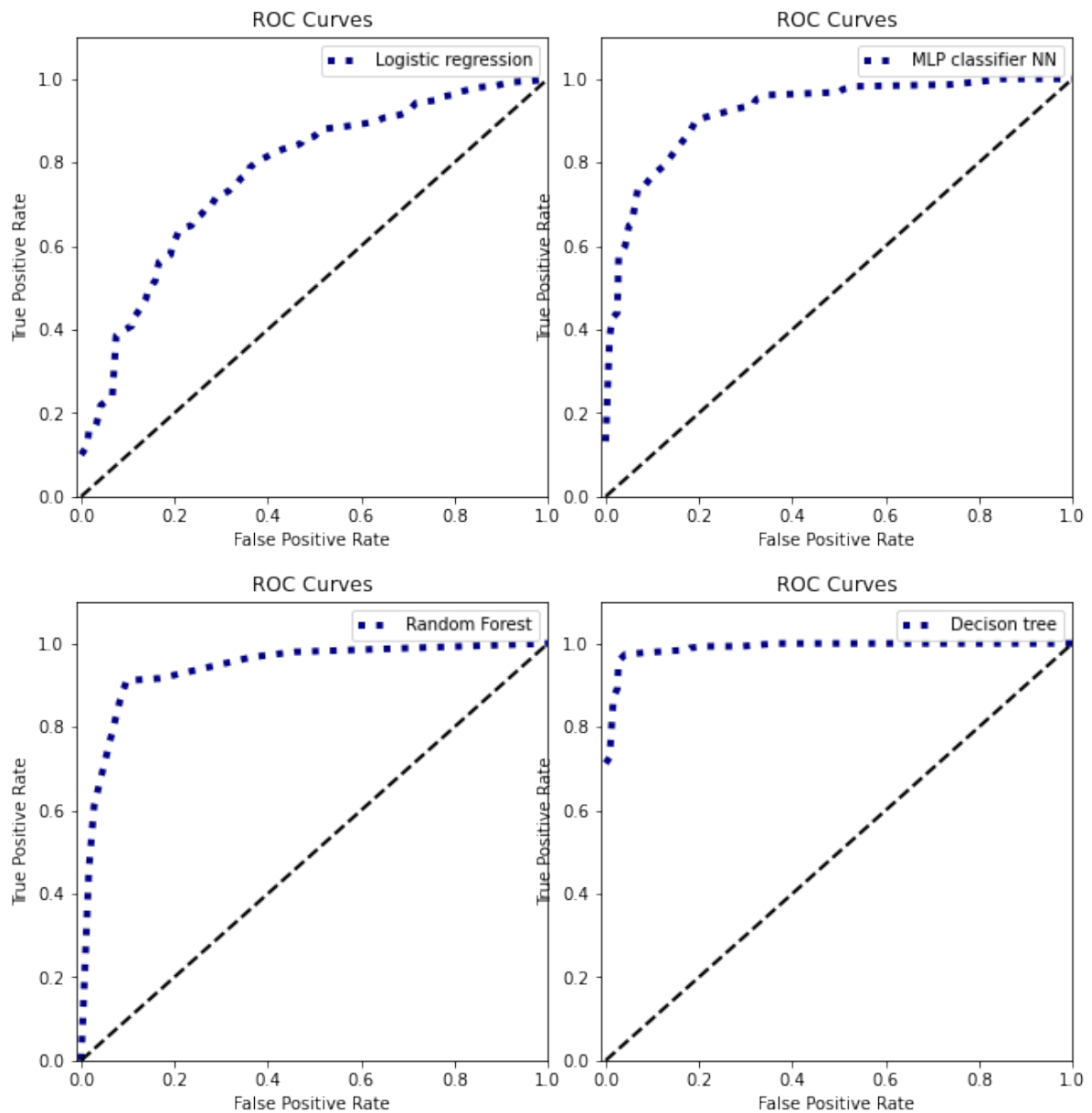
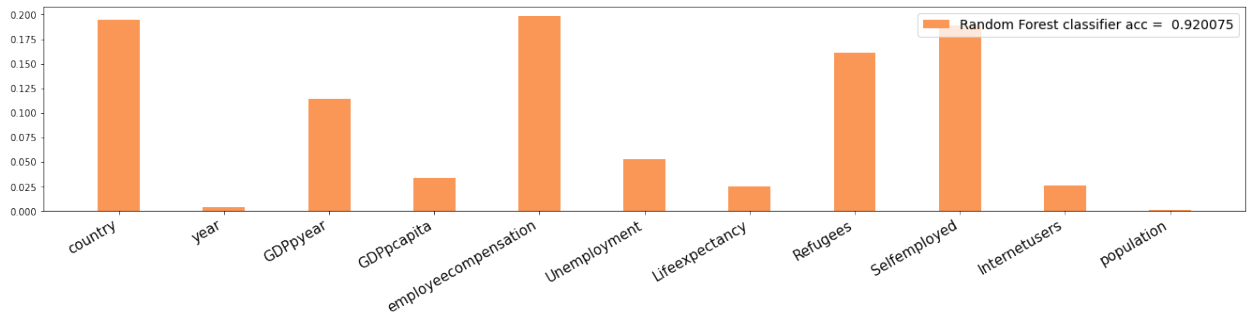


Figure 4.1: Comparaison des performances des modèles selon la courbe ROC

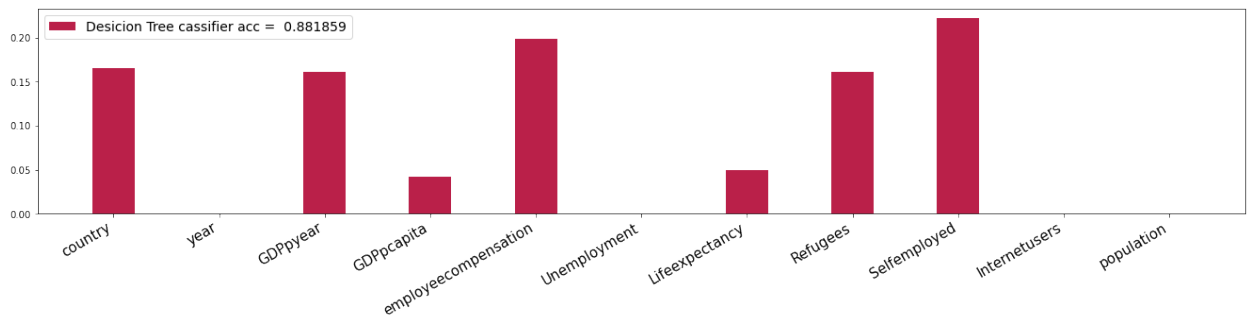
4.1 L'importance des features par modèles de classification :

On a pris les trois modèles les plus performants et pour chacun d'eux on regardait les poids des features qui a permet de séparer les classes :



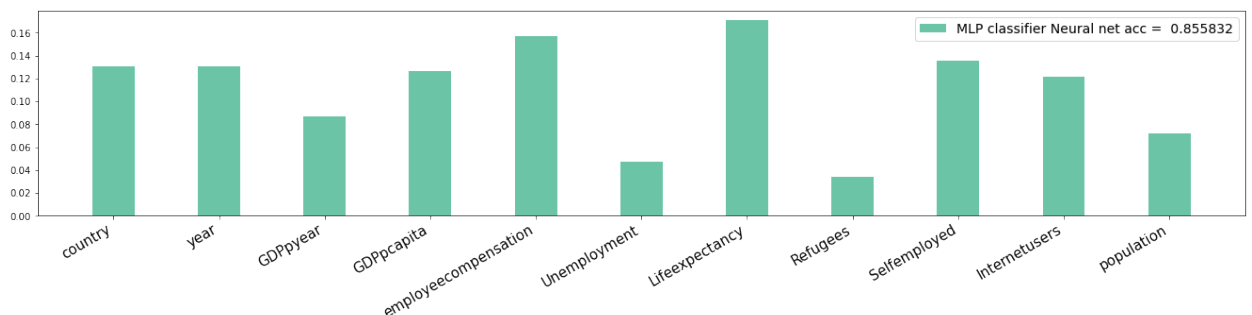
Analyse des features les plus importantes pour le 1er classifieur performant:

- Employee Compensation : la rémunération des employés
- Self-Employed : L'emploi indépendants des travailleurs
- Réfugiés



Les features les plus importantes pour le 2ème classifieur performant :

- La rémunération des employés
- Les travailleurs indépendants
- Réfugiés
- Sigle de produit intérieur brut par année.



Les features les plus importantes pour le 3ème classifieur performant :

C'est les presque les memes facteurs qu'avant mais cette fois y a les features de l'année et des utilisateurs d'internet qui entre en jeux .

5 Synthèses

Après les différentes analyses faites et les résultats des modèles de machine learning montrés on pourra résumer le lien entre le grand taux de suicide et sa situation socio-économique :

5.1 Les facteurs socio-économique favorisant les suicides :

D'après l'étude sur les données de suicide collectées on peut rédiger les points d'analyse suivant :

- L'injustice envers les travailleurs est l'un des facteurs sociaux favorisant le suicide ce qui peut s'expliquer par le grand nombre de décès par ce fait dans les pays à régime autoritaire et autocratique (comme la Russie et la République de Corée figurant parmi top 5 pays des suicides dans le monde) .
- La non indépendance des travailleurs , la soumission et l'exploitation de ces derniers .
- Le grand taux de réfugiés dans le pays, ce qui peut s'expliquer que ces personnes en elles-mêmes qui commettent l'acte suicidaire .
- La pauvreté : le Gdp-par-capita est l'un des facteurs montrés dans nos corrélations avec le suicide .

5.2 Prévention

- Défense des droits des travailleurs (Renoncer à l'exploitation des ouvriers , faire des organisations permettant de surveiller ce processus surtout dans les pays à régime autoritaire et autocratique comme la Russie et la Corée du Nord)
- Intégration de la prévention de suicide dans la politique santé publique de chaque pays(selon notre analyse y a plusieurs pays qui ne disposent pas assez de données à propos de ce fait tellement le sujet est très tabou)
- Traiter le problème des réfugiés (à chaque pays de trouver sa propre solution: les intégrer dans la société par exemple , ...)
- Diminuer le taux de chômage dans les pays pauvres