

Projet d'Apprentissage automatique : ADDICTEST

Filière : Ingénierie Numérique en Data Science et Intelligence Artificielle

Etablissement : École Nationale supérieure d'Arts et Métiers de Rabat

Élément de module : Apprentissage automatique 1

Réalisé par :

**LAKHL Rachid
IDELHAJ Amine
FATIH SALAH EDDINE
SOUAD AYOUB**

Encadré par :

Pr. ALQADI Abderrahim

Remerciements

Ce travail était effectué au sein de l'*École Nationale supérieure d'Arts et Métiers de Rabat* sous l'encadrement de Professeur ALQADI Abderrahim.

On commence par présenter notre plus vive gratitude à notre encadrant de notre projet de fin d'étude. Mr. ALQADI Abderrahim, Professeur de l'élément du module (Apprentissage automatique). Grâce à ses encouragements, sa pédagogie et ses précieux conseils, il a su nous guider pour mener à bien notre projet. On voudrait exprimer ici notre profonde gratitude à son égard et l'estime respectueuse qu'on lui porte.

On voudrait remercier, toute l'équipe pédagogique et administrative d'École Nationale supérieure d'Arts et Métiers de Rabat, qui nous a aidé par ses qualités humaines, ses rigueurs professionnelles.

On remercie les enseignants chercheurs pour leurs efforts de transmettre aux étudiants leurs connaissances professionnelles.

TABLE DES MATIERES

Contents

Remerciements	2
TABLE DES MATIERES	3
Introduction Générale	4
Partie 1: Description du projet et méthode de travail - SCRUM	5
I. Description du projet	5
II. Méthode de travail - SCRUM	5
Partie 2 : Modèle d'apprentissage automatique	7
Partie 3 : Implémentation–Application Django (AddicTest)	21
Conclusion Générale	26

Introduction Générale

Le besoin d'optimiser, planifier ou prendre des décisions en temps réel est partout, même dans notre vie de tous les jours. A chaque instant et face à toute situation, nous sommes dans l'obligation de prendre une décision parmi plusieurs. Le problème, c'est que parfois notre décision dépend d'une multitude de paramètres et d'inconvénients, ce qui rend la vérification d'un choix très difficile.

En prenant l'exemple de secteur de santé, de nombreuses personnes mortes à cause des mauvaises décisions et diagnostic de médecins sur les statuts des personnes, voire même qu'il est été classé comme une des sources principales des décès dans le monde.

Ce projet, qui s'aligne parmi les recherches visant à traiter ce type de problèmes complexes, à pour le but de simuler le réel par un système de prédiction des addictions probable, plus précisément, dans le monde d'Intelligence artificielle, afin de pouvoir diminuer le nombre des décès.

Nous allons présenter au début une vision globale sur notre problématique et le contexte général du projet, puis on va parler sur la méthode du travail SCRUM, ensuite on va entamer la partie de l'apprentissage de données, et à la fin on va présenter la partie réalisation dans laquelle on va présenter les résultats expérimentaux de notre projet.

Partie 1: Description du projet et méthode de travail - SCRUM

I. Description du projet

ADDICTEST est une solution qui vise à relever les nouveaux défis du secteur de la santé et à répondre aux besoins des étudiants universitaires qui ont besoin de prévention de l'addiction. Ainsi, la solution **ADDICTEST** prédit le risque de souffrir d'une addiction à de nombreuses substances dangereuses en se basant sur des modèles avancés d'apprentissage automatique.






II. Méthode de travail - SCRUM

Afin de dérouler le projet dans des conditions standards qui s'inspirent des méthodologies de travail les plus efficaces et dans le cadre de mon équipe de travail nous avons choisi la méthodologie Scrum pour la gestion du projet. Dans un premier temps cette méthodologie favorise la productivité au sein d'une équipe de développement en travaillant sur des objectifs prioritaires et à court terme et en suivant un développement itératif et incrémental avec une planification évolutive basée sur la division de l'ensemble des tâches selon des unités qu'on appelle sprint.

Préparation des sprints

Sprint 1

- Liste des tâches à réaliser :

<input checked="" type="checkbox"/>  MLP-1 Il faut faire une approche analytique et définir les exigences en matière de données afin de comprendre la problématique
<input checked="" type="checkbox"/>  MLP-2 Il faut collecter, comprendre et préparer le jeu de données afin de comprendre la base de données
<input checked="" type="checkbox"/>  MLP-3 Il faut faire une formation Django afin d'être capable de créer une interface web pour le projet
<input checked="" type="checkbox"/>  MLP-4 En tant que programmeur, je veux nettoyer le jeu de donnée afin de la manipuler (Data Cleaning)
<input checked="" type="checkbox"/>  MLP-5 En tant que programmeur, je veux faire un encodage des données pour faire des calculs dans le modele (Data Encoding)

- Avancement du projet (Road Map)

MLP board

Sprint 1

réaliser le modèle d'apprentissage pour la détection de maladie

QUICK FILTERS: Only My Issues Recently Updated

2 days remaining

Complete sprint

Board

TO DO

MLP-5
En tant que programmeur, je veux faire un encodage des données pour faire des calculs dans le modele (Data Encoding)
☒ 

IN PROGRESS

MLP-4
En tant que programmeur, je veux nettoyer le jeu de donnée afin de la manipuler (Data Cleaning)
☒ 

MLP-3
Il faut faire une formation Django afin d'être capable de créer une interface web pour le projet
☒ 

DONE

MLP-1
Il faut faire une approche analytique et définir les exigences en matière de données afin de comprendre la problématique
☒ 

MLP-2
Il faut collecter, comprendre et préparer le jeu de données afin de comprendre la base de données
☒ 

PLAN:

plan



Give feedback

0 warnings

Auto-schedule

Share

Exp

Roadmap Teams Releases Dependencies report

Hierarchy: Story to Story Filters

3M 1Y Fit 15/Nov/21 – 23/Nov/21

SCOPE		Fields			
#	Issue	Target start	Target end	Status	
1	MLP-1 Il faut faire une approche analytique et définir les exigences en matière de ...		15/Nov/21	DONE	
2	MLP-4 En tant que programmeur, je veux nettoyer le jeu de donnée afin de la man...	15/Nov/21	16/Nov/21	IN PROGRESS	
3	MLP-2 Il faut collecter, comprendre et préparer le jeu de données afin de compren...	15/Nov/21	22/Nov/21	IN PROGRESS	
4	MLP-3 Il faut faire une formation Django afin d'être capable de créer une interface...		18/Nov/21	DONE	
5	MLP-5 En tant que programmeur, je veux faire un encodage des données pour fair...	21/Nov/21	23/Nov/21	TO DO	

Sprint 2

Liste des tâches à réaliser :

- MLP-1 En tant que analysteur, je veux analyser les drogues légales pour predira le niveau de consommation des drogues
- MLP-2 En tant que analysteur, je veux analyser toutes les drogues illégales pour predira le niveau de consommation de chaque drogue
- MLP-3 En tant que programmeur, je veux choisir et programmer l'algorithme du machine Learning afin de travailler sur le modelé de prédictions
- MLP-4 En tant que programmeur, je veux préparer un projet Django avec toutes les interfaces nécessaires pour obtenir une application
- MLP-5 En tant que programmeur, je veux integrer le modele du ML dans le projet Django pour obtenir une application complète
- MLP-6 En tant que utilisateur, je veux entrer toutes les informations nécessaires afin de simuler mon risque de développer une addiction
- MLP-7 En tant qu'administrateur, je veux authentifier à mon espace pour voir tous les résultats des utilisateurs

Avancement du projet (Road Map)

MLP board

Sprint 2

réaliser le modèle d'apprentissage pour la détection de maladie

3 days remaining

Complete sprint

Board



QUICK FILTERS: Only My Issues Recently Updated

TO DO

MLP-5
En tant que programmeur, je veux integrer le modele du ML dans le projet Django pour obtenir une application complète



MLP-6
En tant que utilisateur, je veux entrer toutes les informations nécessaires afin de simuler mon risque de développer une addiction



MLP-8
En tant qu'administrateur, je veux authentifier à mon espace pour voir tous les résultats des utilisateurs



IN PROGRESS

MLP-2
En tant que analysteur, je veux analyser toutes les drogues illégales pour predira le niveau de consommation de chaque drogue



MLP-3
En tant que programmeur, je veux choisir et programmer l'algorithme du machine Learning afin de travailler sur le modelé de prédictions



DONE

MLP-1
En tant que analysteur, je veux analyser les drogues légales pour predira le niveau de consommation des drogues



MLP-4
En tant que programmeur, je veux préparer un projet Django avec toutes les interfaces nécessaires pour obtenir une application



Partie 2 : Modèle d'apprentissage automatique

I. Analyse des données

La base de données contient les enregistrements de 1885 personnes. Pour chaque personne, 12 attributs sont connus :

Mesures de la personnalité : NEO-FFI-R (névrosisme, extraversion, ouverture à l'expérience, amabilité et conscience), BIS-11 (impulsivité) et ImpSS (recherche de sensations)

Informations Personnelles : niveau d'éducation, l'âge, le sexe, le pays de résidence et l'origine ethnique.

De plus, les participants ont été interrogés sur leur consommation de 18 drogues légales et illégales.

Drogues Légales : alcool, benzodiazépine, nicotine, caféine, chocolat.

Drogues Illégales : amphétamines, nitrite d'amyle, cannabis, cocaïne, crack, ecstasy, héroïne, kétamine, legalh, LSD, méthadone, Mushroom, l'abus de substances volatiles.

Droque fictif : Semeron - qui a été introduit pour identifier les sur-réclamants.

Pour chaque drogue, ils doivent sélectionner l'une des réponses : n'a jamais utilisé, l'a utilisé il y a plus d'une décennie ou au cours de la dernière décennie, année, mois, semaine ou jour.

II. Description des Attributs

Feature	Description	
Informations de BD	ID	Nombre de l'enregistrement dans la base de données originale
	Age	Age du participant
	Gender	Male ou Female
Informations Personnel	Education	Niveau d'éducation du participant
	Country	Pay d'origine du participant
	Ethnicity	Étnicité du participant
Psychologie	Nscore	Neuroticism [NEO-FFI-R]
	Escore	Extraversion [NEO-FFI-R]
	Oscore	Openness to experience [NEO-FFI-R]
	Ascore	Agreeableness [NEO-FFI-R]
	Cscore	Conscientiousness [NEO-FFI-R]
	Impulsive	impulsiveness [BIS-11]
	SS	Sensation seeing [ImpSS]
	Alcohol	Consommation d'alcool
	Amphet	Consommation d'Amphet
	Amyl	Consommation d'Amyl
	Benzos	Consommation de Benzos
	Caff	Consommation de Caffeine
	Cannabis	Consommation de Cannabis

Drogues	Choc	Consommation de Chocolat
	Coke	Consommation de Cocaine
	Crack	Consommation de Crack
	Ecstasy	Consommation d'Ecstasy
	Heroin	Consommation de Heroin
	Ketamine	Consommation de Ketamine
	Legalh	Consommation de Legalh
	LSD	Consommation d'LSD
	Meth	Consommation de Meth
	Mushroom	Consommation de Mushroom
	Nicotine	Consommation de Nicotine
	VSA	Abus de consommation de substances volatiles
	Semer	Consommation de Semer [Droque fictif, afin de détecter les sur-réclamants]

Désignations des labels de consommation des drogues

CL0 : jamais utilisé

CL1 : utilisé il y a plus d'une décennie

CL2 : utilisé au cours de la dernière décennie

CL3 : utilisé au cours de l'année dernière

CL4 : utilisé au cours du mois dernier

CL5 : utilisé au cours de la dernière semaine

CL6 : utilisé au cours du dernier jour

III. Data Cleaning

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

pd.set_option('max_columns', None) # Afficher tous les colonnes
sns.set(style="white", color_codes=True) # utiliser le style de sns

data = pd.read_csv('data/Drug_Consumption.csv')
data = data.drop('ID', axis=1) # Supprimer la colonne "ID" parcequ'il n'as pas de valeur pour l'analyse
data.head()
```

	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	AScore	Cscore	Impulsive	SS	Alcohol	Amphet	Amyl	Benzos	Caff	Cannabis	Choc	Coke	Crack	Ecstasy	Heroin	Ketamine
0	25-34	M	Doctorate degree	UK	White	-0.67825	1.93886	1.43533	0.76096	-0.14277	-0.71126	-0.21575	CL5	CL2	CL2	CL0	CL6	CL4	CL6	CL3	CL0	CL4	CL0	CL2
1	35-44	M	Professional certificate/ diploma	UK	White	-0.46725	0.80523	-0.84732	-1.62090	-1.01450	-1.37983	0.40148	CL6	CL0	CL0	CL0	CL6	CL3	CL4	CL0	CL0	CL0	CL0	CL0
2	18-24	F	Masters degree	UK	White	-0.14882	-0.80615	-0.01928	0.59042	0.58489	-1.37983	-1.18084	CL4	CL0	CL0	CL3	CL5	CL2	CL4	CL2	CL0	CL0	CL0	CL2
3	35-44	F	Doctorate degree	UK	White	0.73545	-1.63340	-0.45174	-0.30172	1.30612	-0.21712	-0.21575	CL4	CL1	CL1	CL0	CL6	CL3	CL6	CL0	CL0	CL1	CL0	CL0
4	65+	F	Left school at 18 years	Canada	White	-0.67825	-0.30033	-1.55521	2.03972	1.63088	-1.37983	-1.54858	CL2	CL0	CL0	CL0	CL6	CL0	CL4	CL0	CL0	CL0	CL0	CL0


```
# Chercher Les valeurs manquantes
data.isna().sum().sum()
```


La base de données ne contient aucune valeurs manquante

```
print("Nombre d'enregistrement original : ", data.shape[0])
```

Nombre d'enregistrement original : 1884

```
print("Nombre d'enregistrements non voulu : ", data[data['Semer']!= 'CL0'].shape[0])
```

Nombre d'enregistrements non voulu : 8

```
# Supprimer la colonne et les entrées des personnes ayant dit qu'il ont consommé la drogue fictive "Semer"
data = data.drop(data[data['Semer'] != 'CL0'].index)
data = data.drop('Semer', axis=1)
data = data.drop('Choc', axis=1)
data = data.drop('Caff', axis=1)
data = data.drop('Benzos', axis=1)
data.reset_index()
print("Nombre d'enregistrement sans la drogue fictive : ", data.shape[0])
```

Nombre d'enregistrement sans la drogue fictif : 1876

IV. Data Encoding

```
data['Gender'] = data['Gender'].map({'M':1, 'F':0})
```

```
ordinal_features = ['Age', 'Education', 'Alcohol', 'Amyl', 'Amphet', 'Cannabis', 'Coke', 'Crack',
                   'Ecstasy', 'Heroin', 'Ketamine', 'Legalh', 'LSD', 'Meth', 'Mushrooms', 'Nicotine', 'VSA']
ordinal_orderings = [
    ['18-24', '25-34', '35-44', '45-54', '55-64', '65+'],
    ['Left school before 16 years', 'Left school at 16 years', 'Left school at 17 years', 'Left school at 18 years',
     'Some college or university, no certificate or degree',
     'Professional certificate/ diploma', 'University degree', 'Masters degree', 'Doctorate degree'],
    ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'],
    ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'],
    ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'],
    ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6'], ['CL0', 'CL1', 'CL2', 'CL3', 'CL4', 'CL5', 'CL6']
]
nominal_features = ['Country', 'Ethnicity']
```

```
# Ordinal encoding
def ordinal_encoder(df, columns, oredrings):
    df = df.copy()
    for column, ordering in zip(columns, oredrings):
        df[column] = df[column].apply(lambda x : ordering.index(x))
    return df

# nominal/categorical encoding
def cat_converter(df, columns):
    df = df.copy()
    for column in columns:
        df[column] = df[column].astype('category').cat.codes
    return df
```

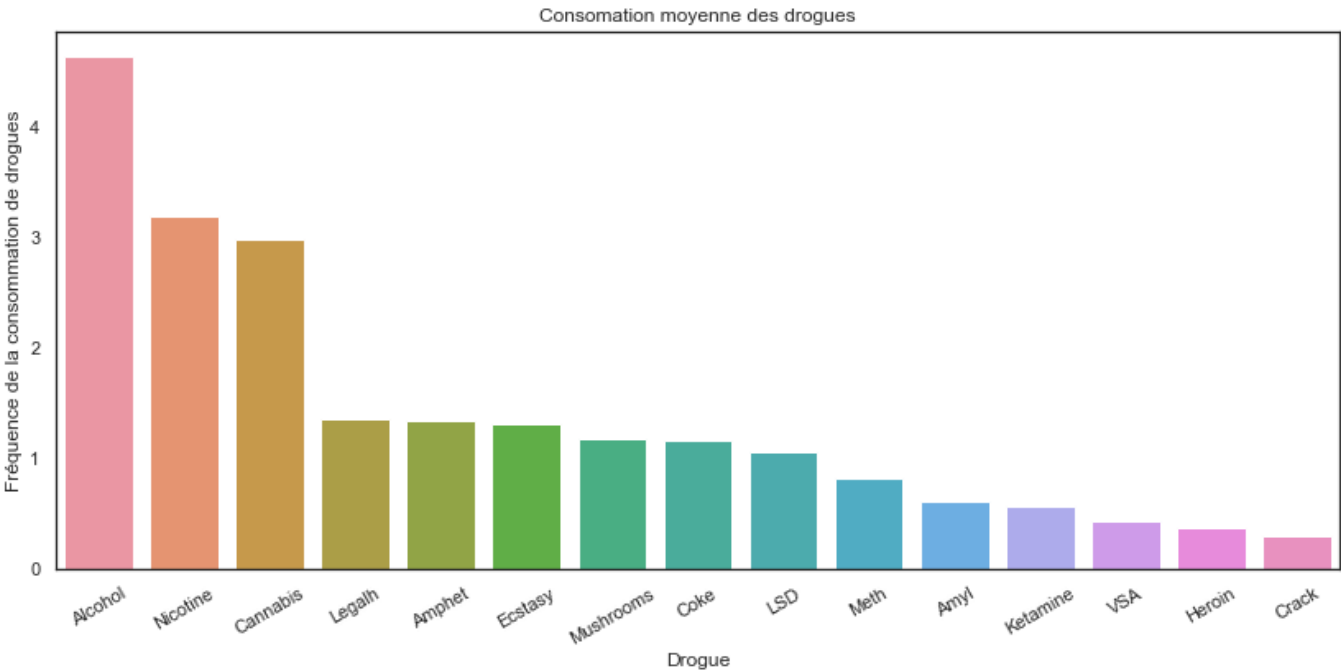
```
data = ordinal_encoder(data, ordinal_features, ordinal_orderings)
data = cat_converter(data, nominal_features)
data.head()
```

```
data = data.astype('float64')
data.head()
```

➔ Les données sont devenues des données numériques

V. Analyse générale des données

data.describe()													
	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	AScore	Cscore	Impulsive	SS	Alcoh
count	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.000000	1876.0000
mean	1.349680	0.501066	4.995203	4.825160	5.742004	-0.000718	-0.001646	-0.002915	-0.000169	-0.000391	0.005412	-0.006782	4.6375
std	1.278042	0.500132	1.764684	1.409047	1.021667	0.998682	0.997596	0.995866	0.996730	0.997923	0.954389	0.961949	1.3285
min	0.000000	0.000000	0.000000	0.000000	0.000000	-3.464360	-3.273930	-3.273930	-3.464360	-3.464360	-2.555240	-2.078480	0.0000
25%	0.000000	0.000000	4.000000	5.000000	6.000000	-0.678250	-0.695090	-0.717270	-0.606330	-0.652530	-0.711260	-0.525930	4.0000
50%	1.000000	1.000000	5.000000	5.000000	6.000000	0.042570	0.003320	-0.019280	-0.017290	-0.006650	-0.217120	0.079870	5.0000
75%	2.000000	1.000000	6.000000	6.000000	6.000000	0.629670	0.637790	0.723300	0.760960	0.628243	0.529750	0.765400	6.0000
max	5.000000	1.000000	8.000000	6.000000	6.000000	3.273930	3.273930	2.901610	3.464360	3.464360	2.901610	1.921730	6.0000



- Comme prévu on constate que l'Alcool, Nicotine et Cannabis sont les drogues les plus consommées.

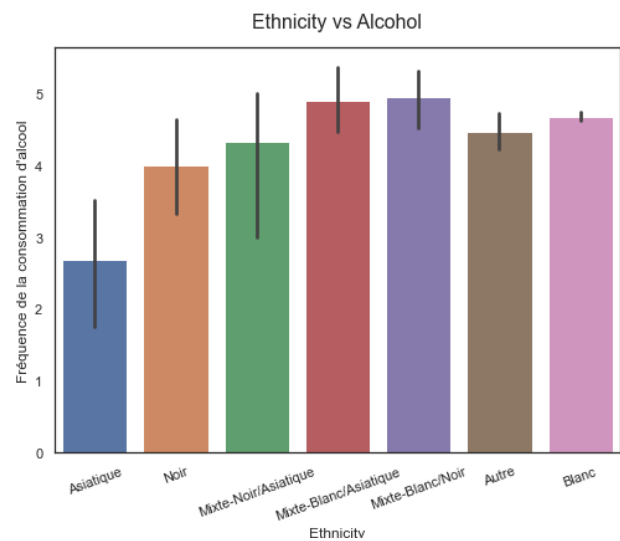
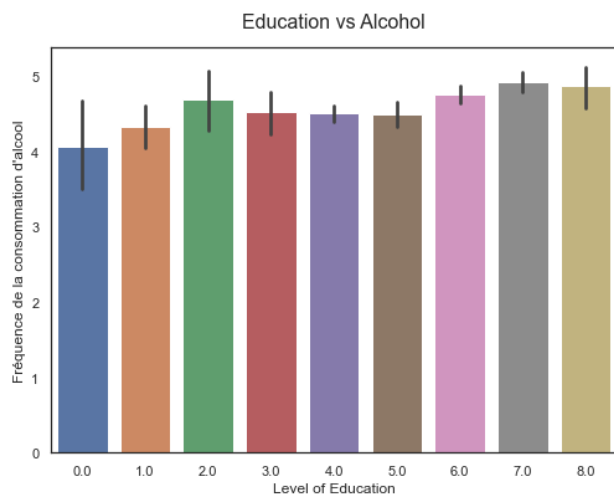
VI. Analyse des drogues légales



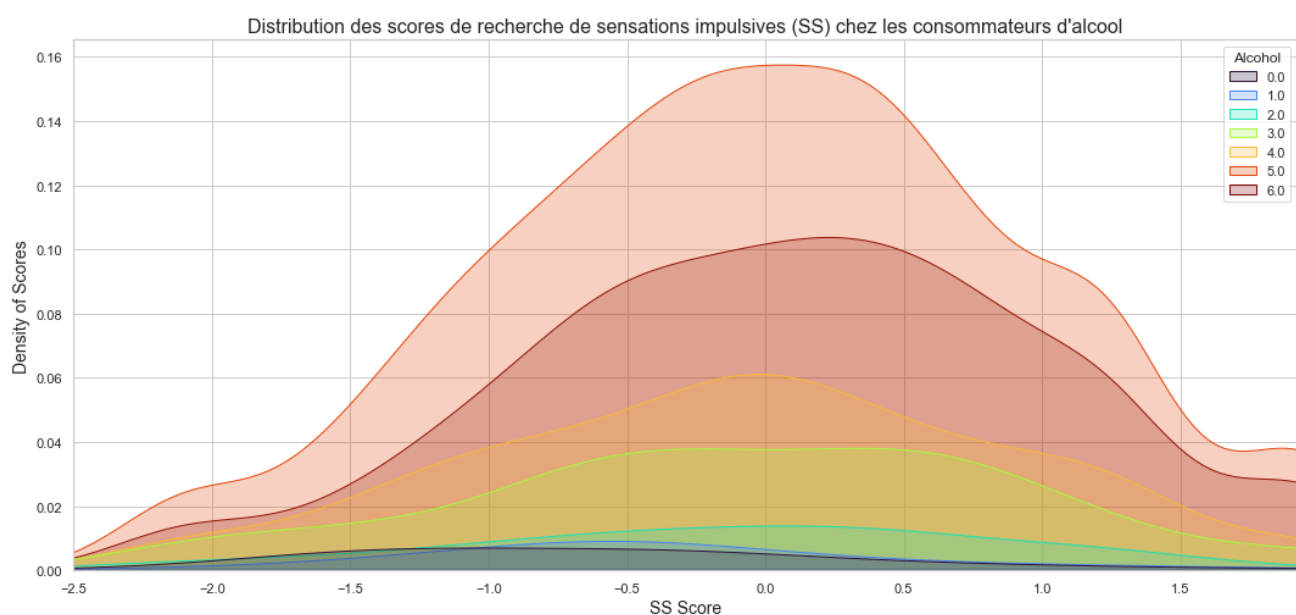
Conclusion

- Une légère corrélation positive entre la Nicotine et le score de Recherche de Sensations (SS), ainsi qu'une corrélation un peu moins prononcée avec l'Impulsivité, Oscore ainsi que le Sexe.
- Une légère corrélation négative entre la Nicotine et l'Age, l'Education et le Cscore
- Une faible corrélation positive entre l'Alcool et Recherche de Sensations (SS), ainsi que l'Ethnicité et l'Éducation.
- Le niveau des corrélations entre les informations personnelles et l'Alcool sont moins prononcées que leurs niveaux de corrélation avec la Nicotine.

VII. Analyse de L'Alcool



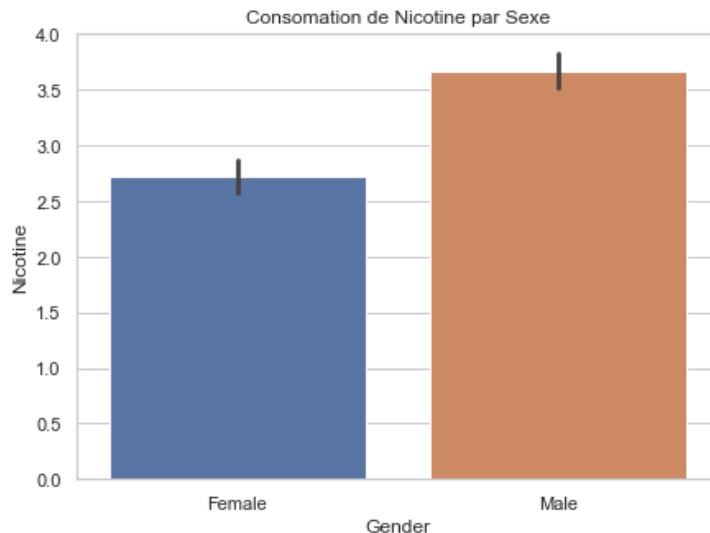
Conclusion : La consommation d'alcool montre une faible corrélation positive avec le niveau d'éducation, de sorte que le plus haut le niveau d'enseignement est associé à une consommation d'alcool plus fréquente. Plus précisément, ceux qui ont un diplôme universitaire ou supérieur (c'est-à-dire un Master ou un doctorat) ont consommé le plus d'alcool.



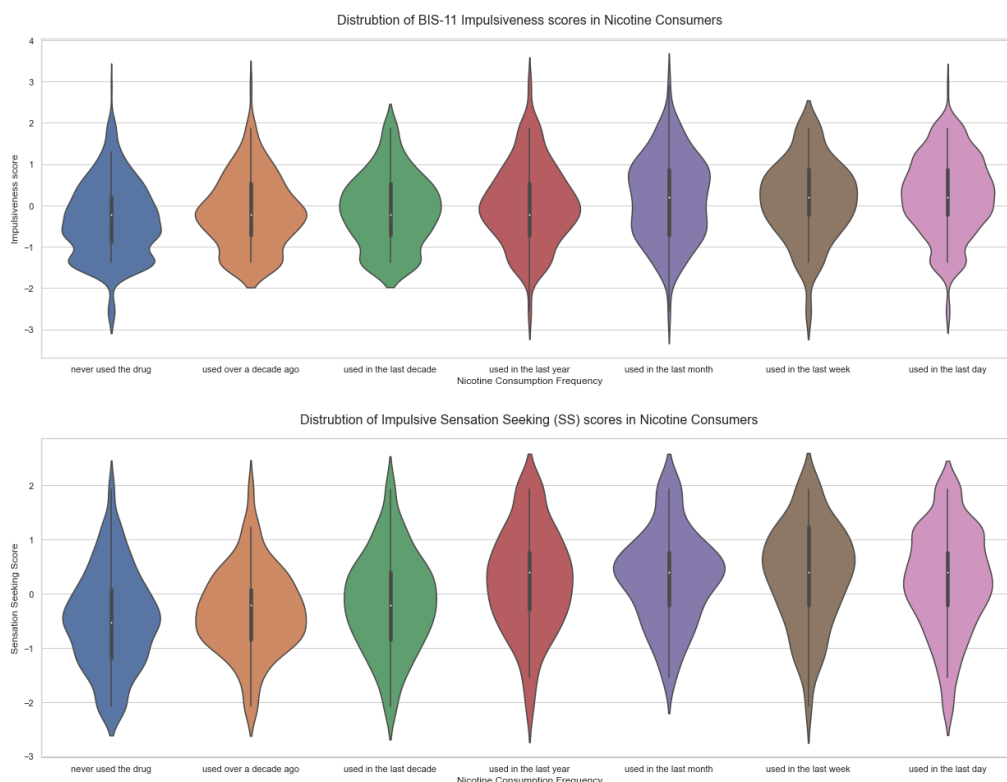
Conclusion :

Les consommateurs fréquents d'alcool ont tendance à afficher des scores de recherche de sensations impulsives (SS) plus élevés, de sorte qu'une majorité de non-consommateurs ("jamais bu" ou "consommé il y a plus de dix ans") avaient un score SS proche de -1,0 tandis que les consommateurs fréquents ("consommé il y a une semaine" ou "un jour") avaient tendance à obtenir un score compris entre 0 et 0,25.

VIII. Analyse de Nicotine

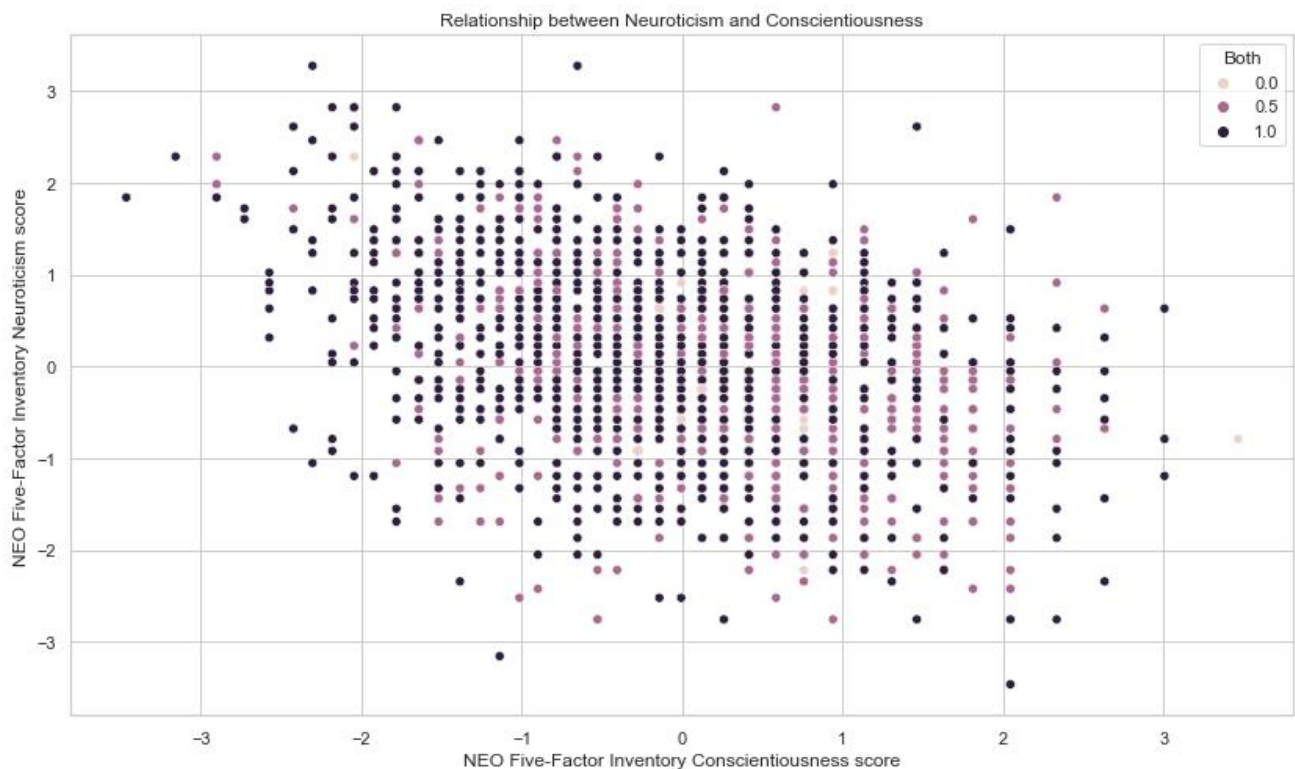
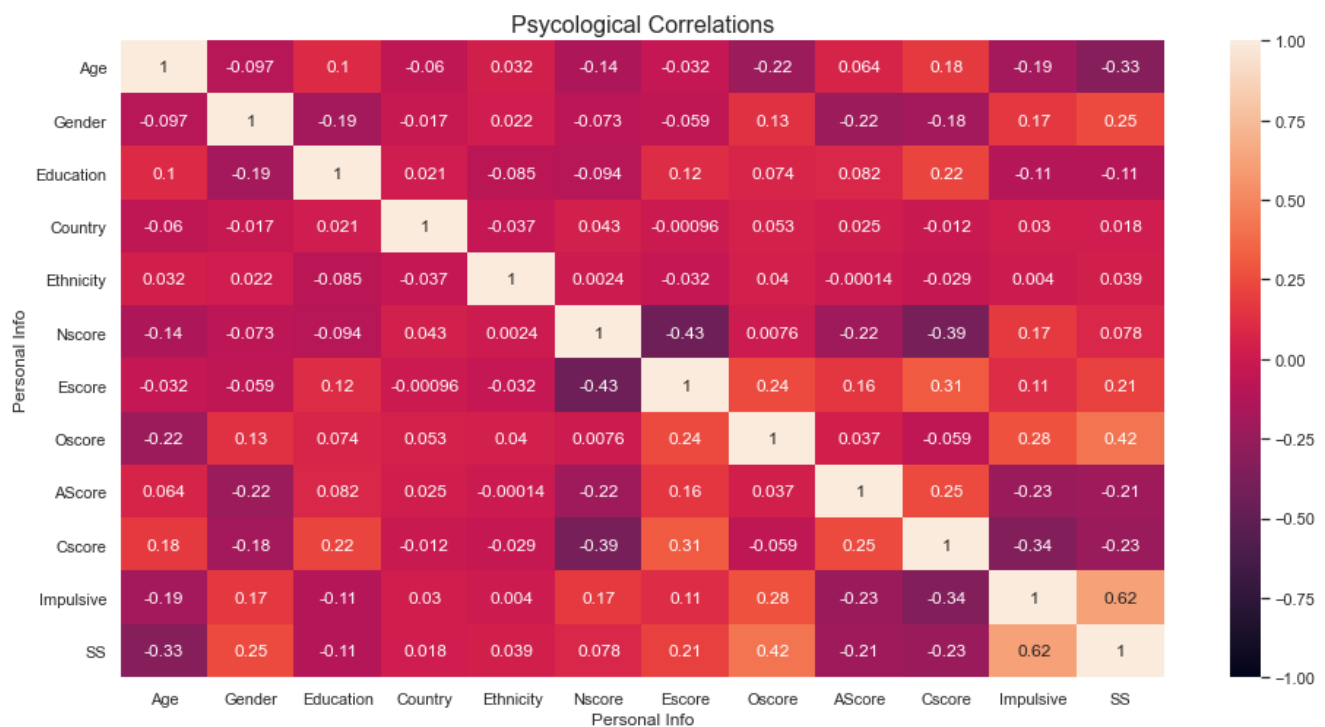


Conclusion : Contrairement à l'alcool, la nicotine a montré une différence marquée dans la consommation entre les sexes, les hommes consommant plus de nicotine que les femmes.



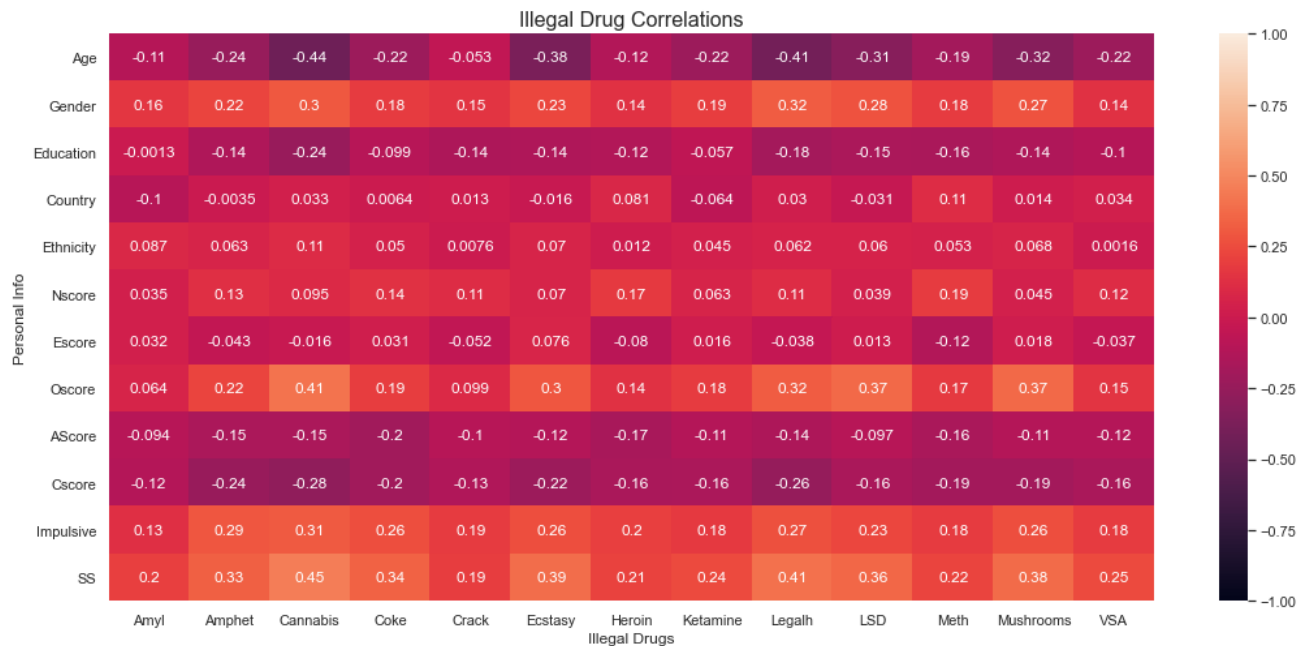
Conclusion : La consommation de nicotine est positivement corrélée à la fois avec le score d'impulsivité BIS-11 et le SS. Cependant, cette relation était légèrement plus prononcée dans le SS.

■ Corrélation entre les scores psychologiques

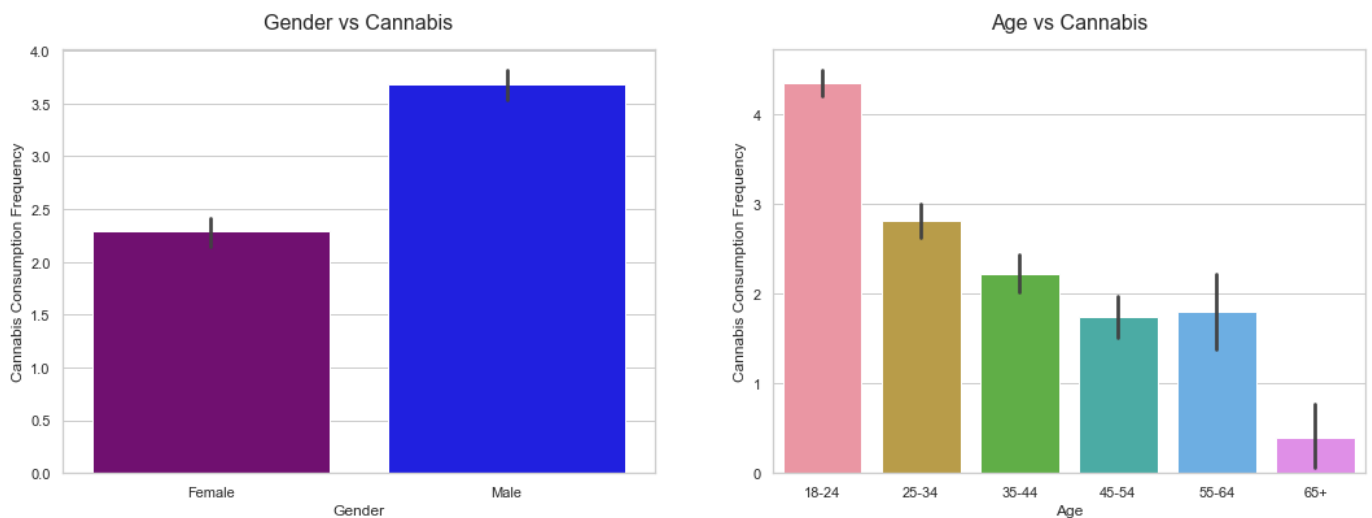


Conclusion : Fait intéressant, on a remarqué une forte corrélation négative entre le Nscore et la Conscience (Cscore). Pour mieux comprendre comment cette relation était liée à la consommation de drogue, j'ai examiné la consommation de nicotine et de l'alcool. Sans surprise, ceux qui consommaient à la fois de la nicotine et de l'alcool avaient tendance à avoir des Nscores plus élevés et des Cscores inférieurs, tandis que ceux qui n'utilisaient aucune drogue affichaient des Cscores élevés et des Nscores faibles.

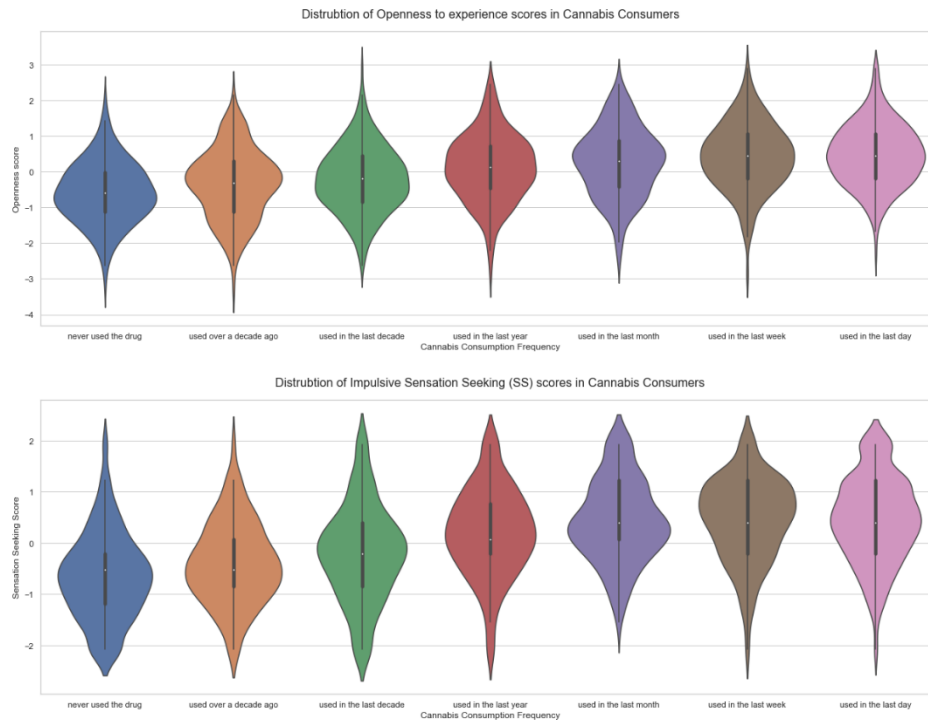
IX. Analyse des drogues illégales



X. Annalyse du Cannabis



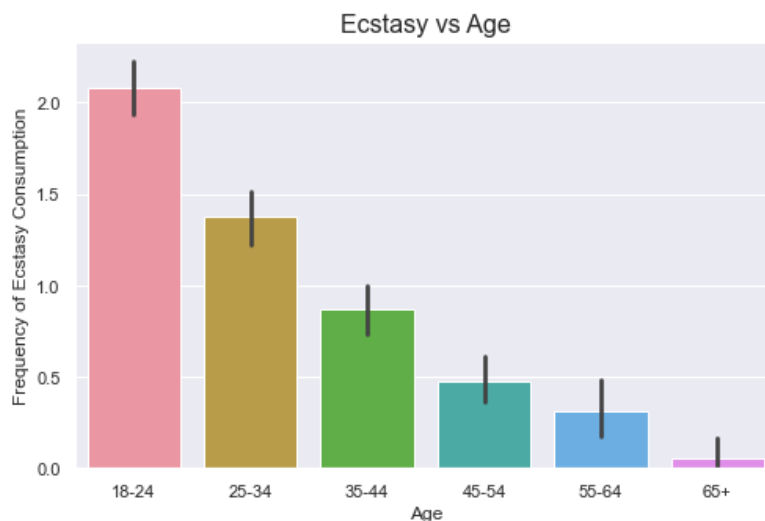
Conclusion : Pour le cannabis, les hommes ont consommé la drogue plus fréquemment que les femmes. De plus, la fréquence de consommation de cannabis était négativement corrélée avec l'âge, de sorte que les individus plus jeunes en consommaient plus fréquemment que les individus plus âgés.



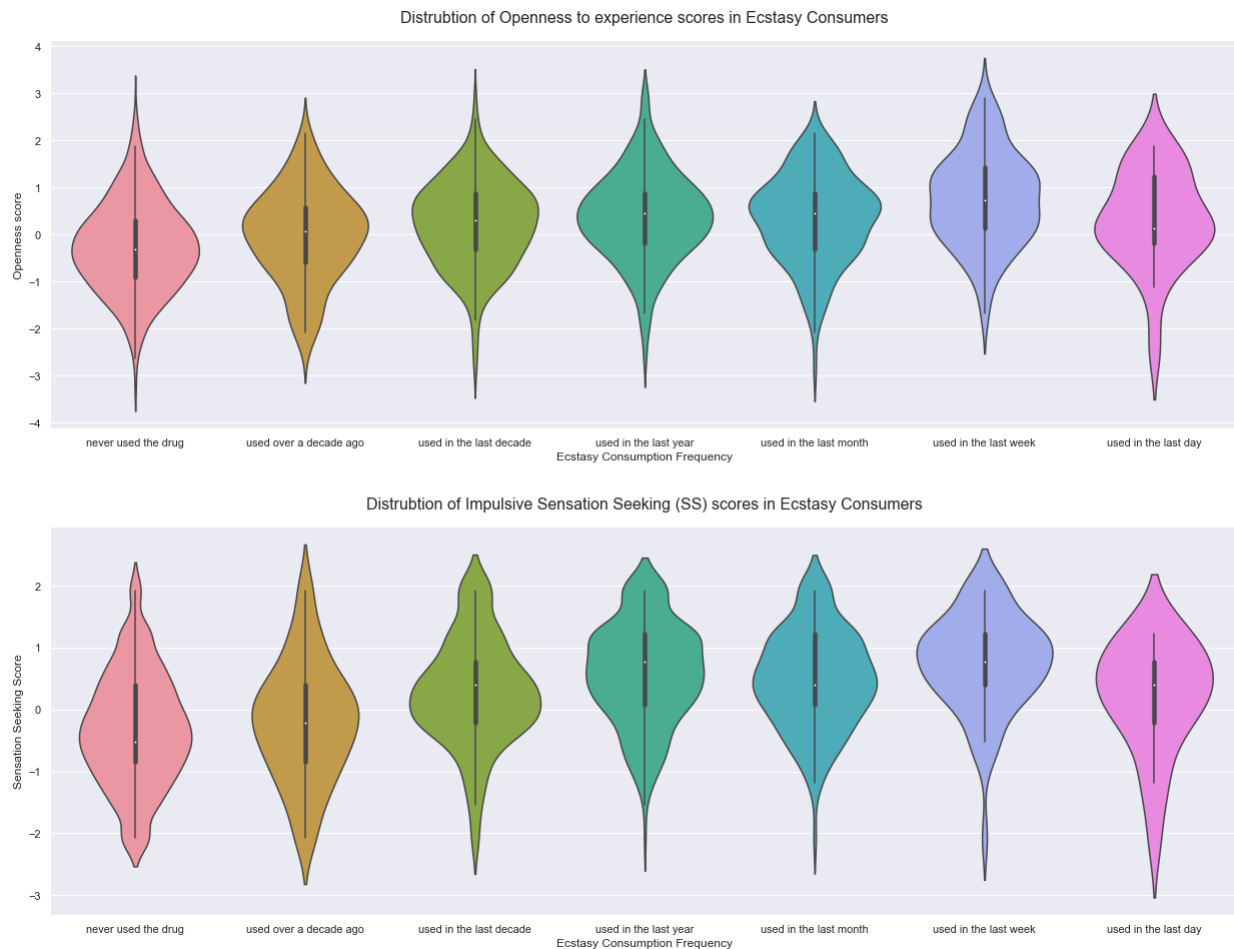
Conclusion : Le Cannabis est positivement corrélé avec le SS. Les personnes qui n'ont jamais ou rarement consommé l'une des drogues illégales ont montré des scores SS de -0,5 à -1,0, tandis que les scores des consommateurs fréquents variaient entre 0,5 et 1,5.

Le cannabis a aussi montré une corrélation positive avec l'ouverture à l'expérience (Oscore), les scores de la plupart des consommateurs occasionnels et peu fréquents allant de 0 à -0,5 tandis que les consommateurs les plus fréquents avaient des scores compris entre 0,5 et 1,0.

XI. Analyse de l'Ecstasy



Pour l'Ecstasy, la fréquence de consommation était négativement corrélée avec l'âge, de sorte que les individus plus jeunes en consommaient plus fréquemment que les individus plus âgés.

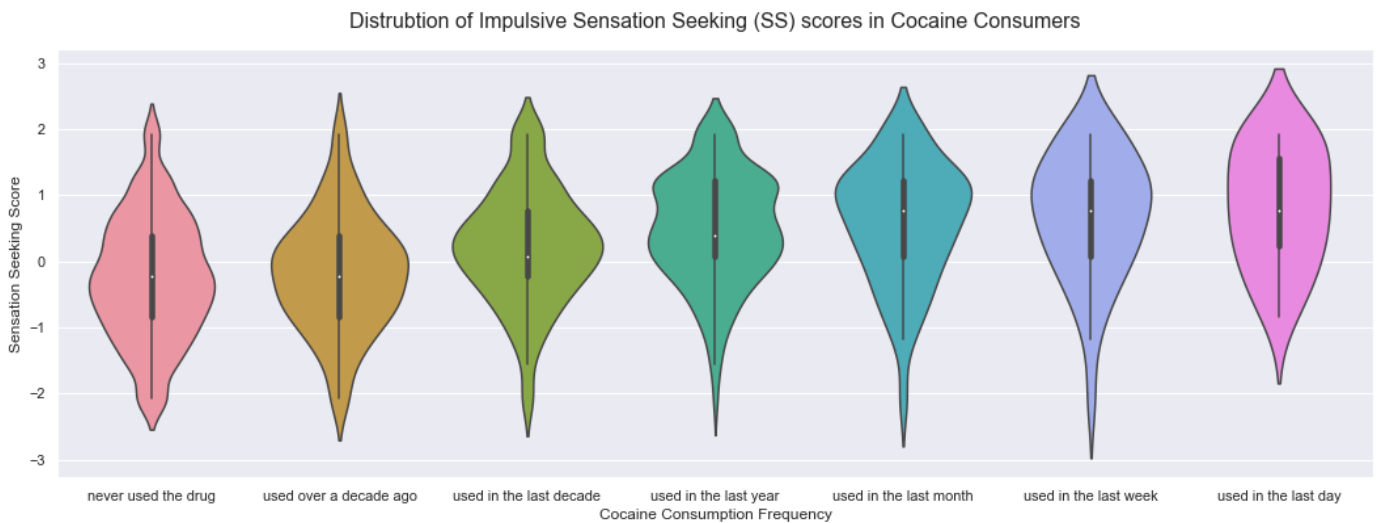


Conclusion :

L'Ecstasy est positivement corrélé avec le SS. Les personnes qui n'ont jamais ou rarement consommé l'une des drogues illégales ont montré des scores SS de -0,5 à -1,0, tandis que les scores des consommateurs fréquents variaient entre 0,5 et 1,5.

L'Ecstasy a également montré une corrélation positive avec l'ouverture à l'expérience (Oscore), les scores de la plupart des consommateurs occasionnels et peu fréquents allant de 0 à -0,5 tandis que les consommateurs les plus fréquents avaient des scores compris entre 0,5 et 1,0.

XII. Analyse de la Cocaine



Conclusion :

La Cocaine est positivement corrélée avec le SS. Les personnes qui n'ont jamais ou rarement consommé l'une des drogues illégales ont montré des scores SS de -0,5 à -1,0, tandis que les scores des consommateurs fréquents variaient entre 0,5 et 2.

XIII. Conclusions Finales sur l'analyse de données :

Étant donné que le SS semblait avoir la corrélation positive la plus élevée pour toutes les drogues (légales ou illégales), en tant que mesure de la personnalité, il peut s'agir de la mesure la plus robuste pour l'usage et/ou l'abus de drogues. Ce score de personnalité peut nous aider à mieux comprendre les personnes les plus à risque de souffrir de toxicomanie ou de consommation de drogues en général, ce qui peut fournir aux médecins et aux thérapeutes une compréhension plus claire et plus nuancée de cette population de patients.

XIV. Prédiction de consommation des drogues

1. Model Predecting

```
def predict_drug_usage(model ,drug):  
    X_train, X_test, y_train, y_test = train_test_split(data[personal_info], data[drug], test_size=0.2, random_state=1)  
    model.fit(X_train, y_train)  
    y_predict = model.predict(X_test)  
    print("Accuracy Score de",["",drug,""] : ",model.score(X_test,y_test))  
    print("F1 Score de",["",drug,""] : ",metrics.f1_score(y_test,y_predict))
```

2. Naive Bayes

Accuracy Score de [Alcohol] : 0.9414893617021277
F1 Score de [Alcohol] : 0.9695290858725762

Accuracy Score de [Nicotine] : 0.6888297872340425
F1 Score de [Nicotine] : 0.7771428571428572

Accuracy Score de [Amyl] : 0.7340425531914894
F1 Score de [Amyl] : 0.23076923076923075

Accuracy Score de [Amphet] : 0.7287234042553191
F1 Score de [Amphet] : 0.6506849315068494

Accuracy Score de [Cannabis] : 0.7686170212765957
F1 Score de [Cannabis] : 0.8277227722722728

Accuracy Score de [Coke] : 0.6914893617021277
F1 Score de [Coke] : 0.5999999999999999

Accuracy Score de [Crack] : 0.8643617021276596
F1 Score de [Crack] : 0.2608695652173913

Accuracy Score de [Ecstasy] : 0.7154255319148937
F1 Score de [Ecstasy] : 0.6603174603174603

Accuracy Score de [Heroin] : 0.848404255319149
F1 Score de [Heroin] : 0.27848101265822783

Accuracy Score de [Ketamine] : 0.7553191489361702
F1 Score de [Ketamine] : 0.3134328358208955

Accuracy Score de [Legalh] : 0.7154255319148937
F1 Score de [Legalh] : 0.6514657980456026

Accuracy Score de [LSD] : 0.726063829787234
F1 Score de [LSD] : 0.5654008438818565

Accuracy Score de [Meth] : 0.7127659574468085
F1 Score de [Meth] : 0.425531914893617

Accuracy Score de [Mushrooms] : 0.6968085106382979
F1 Score de [Mushrooms] : 0.6174496644295302

Accuracy Score de [VSA] : 0.848404255319149
F1 Score de [VSA] : 0.3132530120481928

3. Random Forest Classifier

Accuracy Score de [Alcohol] : 0.9601063829787234
F1 Score de [Alcohol] : 0.9796472184531885

Accuracy Score de [Nicotine] : 0.7021276595744681
F1 Score de [Nicotine] : 0.7902621722846442

Accuracy Score de [Amyl] : 0.8085106382978723
F1 Score de [Amyl] : 0.1

Accuracy Score de [Amphet] : 0.6914893617021277
F1 Score de [Amphet] : 0.532258064516129

Accuracy Score de [Cannabis] : 0.7952127659574468
F1 Score de [Cannabis] : 0.846307385229541

Accuracy Score de [Coke] : 0.6968085106382979
F1 Score de [Coke] : 0.5169491525423728

Accuracy Score de [Crack] : 0.8936170212765957
F1 Score de [Crack] : 0.047619047619047616

Accuracy Score de [Ecstasy] : 0.7127659574468085
F1 Score de [Ecstasy] : 0.6142857142857143

Accuracy Score de [Heroin] : 0.8696808510638298
F1 Score de [Heroin] : 0.0392156862745098

Accuracy Score de [Ketamine] : 0.7925531914893617
F1 Score de [Ketamine] : 0.04878048780487805

Accuracy Score de [Legalh] : 0.7579787234042553
F1 Score de [Legalh] : 0.6829268292682927

Accuracy Score de [LSD] : 0.7712765957446809
F1 Score de [LSD] : 0.5376344086021505

Accuracy Score de [Meth] : 0.7872340425531915
F1 Score de [Meth] : 0.3442622950819672

Accuracy Score de [Mushrooms] : 0.7579787234042553
F1 Score de [Mushrooms] : 0.6513409961685823

Accuracy Score de [VSA] : 0.851063829787234
F1 Score de [VSA] : 0.0

4. SVM Classifier

```
Accuracy Score de [ Alcohol ] : 0.9627659574468085
F1 Score de [ Alcohol ] : 0.9810298102981029
-----
Accuracy Score de [ Nicotine ] : 0.7047872340425532
F1 Score de [ Nicotine ] : 0.8146911519198665
-----
Accuracy Score de [ Amyl ] : 0.824468085106383
F1 Score de [ Amyl ] : 0.0
-----
Accuracy Score de [ Amphet ] : 0.7287234042553191
F1 Score de [ Amphet ] : 0.5641025641025641
-----
Accuracy Score de [ Cannabis ] : 0.7712765957446809
F1 Score de [ Cannabis ] : 0.8371212121212123
-----
Accuracy Score de [ Coke ] : 0.675531914893617
F1 Score de [ Coke ] : 0.37755102040816324
-----
Accuracy Score de [ Crack ] : 0.8909574468085106
F1 Score de [ Crack ] : 0.0
-----
Accuracy Score de [ Ecstasy ] : 0.7420212765957447
F1 Score de [ Ecstasy ] : 0.6472727272727272
```

```
Accuracy Score de [ Heroin ] : 0.8723404255319149
F1 Score de [ Heroin ] : 0.0
-----
Accuracy Score de [ Ketamine ] : 0.8031914893617021
F1 Score de [ Ketamine ] : 0.0
-----
Accuracy Score de [ Legalh ] : 0.7712765957446809
F1 Score de [ Legalh ] : 0.6950354609929078
-----
Accuracy Score de [ LSD ] : 0.7686170212765957
F1 Score de [ LSD ] : 0.47272727272727266
-----
Accuracy Score de [ Meth ] : 0.7925531914893617
F1 Score de [ Meth ] : 0.0
-----
Accuracy Score de [ Mushrooms ] : 0.7792553191489362
F1 Score de [ Mushrooms ] : 0.6795366795366794
-----
Accuracy Score de [ VSA ] : 0.8643617021276596
F1 Score de [ VSA ] : 0.0
```

5. Logistic Regression

```
Accuracy Score de [ Alcohol ] : 0.9601063829787234
F1 Score de [ Alcohol ] : 0.9796472184531885
-----
Accuracy Score de [ Nicotine ] : 0.6968085106382979
F1 Score de [ Nicotine ] : 0.7964285714285714
-----
Accuracy Score de [ Amyl ] : 0.8191489361702128
F1 Score de [ Amyl ] : 0.028571428571428574
-----
Accuracy Score de [ Amphet ] : 0.7101063829787234
F1 Score de [ Amphet ] : 0.5240174672489083
-----
Accuracy Score de [ Cannabis ] : 0.75
F1 Score de [ Cannabis ] : 0.8233082706766918
-----
Accuracy Score de [ Coke ] : 0.7047872340425532
F1 Score de [ Coke ] : 0.4884792626728111
-----
Accuracy Score de [ Crack ] : 0.8909574468085106
F1 Score de [ Crack ] : 0.0
-----
Accuracy Score de [ Ecstasy ] : 0.726063829787234
F1 Score de [ Ecstasy ] : 0.6113207547169812
```

```
Accuracy Score de [ Heroin ] : 0.8696808510638298
F1 Score de [ Heroin ] : 0.0
-----
Accuracy Score de [ Ketamine ] : 0.8058510638297872
F1 Score de [ Ketamine ] : 0.0759493670886076
-----
Accuracy Score de [ Legalh ] : 0.7446808510638298
F1 Score de [ Legalh ] : 0.6571428571428571
-----
Accuracy Score de [ LSD ] : 0.7579787234042553
F1 Score de [ LSD ] : 0.46783625730994155
-----
Accuracy Score de [ Meth ] : 0.7872340425531915
F1 Score de [ Meth ] : 0.2156862745098039
-----
Accuracy Score de [ Mushrooms ] : 0.7127659574468085
F1 Score de [ Mushrooms ] : 0.5304347826086957
-----
Accuracy Score de [ VSA ] : 0.8590425531914894
F1 Score de [ VSA ] : 0.03636363636363636
```

6. Logistic Regression avec GridSearch

```
tuned_parameters = [{"solver" : ['liblinear'], "multi_class" : ['ovr'],
    "class_weight" : [None, 'balanced'], "max_iter" : [100, 200, 500, 800, 1600]},
    {"solver" : ['newton-cg'], "multi_class" : ['ovr', 'multinomial'],
    "class_weight" : [None, 'balanced'], "max_iter" : [100, 200, 500, 800, 1600]}]
model_cv = GridSearchCV(model, tuned_parameters, cv=10, scoring='f1')
for drug in legal_drugs:
    predict_drug_usage(model_cv, drug)
    print('-----')
for drug in illegal_drugs:
    predict_drug_usage(model_cv, drug)
    print('-----')
```

Accuracy Score de [Alcohol] : 0.9796472184531885
F1 Score de [Alcohol] : 0.9796472184531885

Accuracy Score de [Nicotine] : 0.7964285714285714
F1 Score de [Nicotine] : 0.7964285714285714

Accuracy Score de [Amyl] : 0.3700440528634361
F1 Score de [Amyl] : 0.3700440528634361

Accuracy Score de [Amphet] : 0.6449511400651466
F1 Score de [Amphet] : 0.6449511400651466

Accuracy Score de [Cannabis] : 0.8233082706766918
F1 Score de [Cannabis] : 0.8233082706766918

Accuracy Score de [Coke] : 0.6143790849673203
F1 Score de [Coke] : 0.6143790849673203

Accuracy Score de [Crack] : 0.3492063492063492
F1 Score de [Crack] : 0.3492063492063492

Accuracy Score de [Ecstasy] : 0.6624203821656051
F1 Score de [Ecstasy] : 0.6624203821656051

Accuracy Score de [Heroin] : 0.3837209302325581
F1 Score de [Heroin] : 0.3837209302325581

Accuracy Score de [Ketamine] : 0.43555555555555553
F1 Score de [Ketamine] : 0.43555555555555553

Accuracy Score de [Legalh] : 0.6945337620578779
F1 Score de [Legalh] : 0.6945337620578779

Accuracy Score de [LSD] : 0.6067415730337079
F1 Score de [LSD] : 0.6067415730337079

Accuracy Score de [Meth] : 0.4622222222222222
F1 Score de [Meth] : 0.4622222222222222

Accuracy Score de [Mushrooms] : 0.6775244299674267
F1 Score de [Mushrooms] : 0.6775244299674267

Accuracy Score de [VSA] : 0.37288135593220334
F1 Score de [VSA] : 0.37288135593220334

Après l'application de plusieurs modèles (Naïve Bayes, Random Forest Classifieur, SVM Classifieur et régression logistique) et le GridSearch, on a trouvé que le millier résultat de prédiction est donné par le GridSearch sur régression logistique

GridSearch : C'est une méthode d'optimisation (hyperparameter optimization) qui va nous permettre de tester une série de paramètres et de comparer les performances pour en déduire le meilleur paramétrage en utilisant la validation croisée.

Partie 3 : Implémentation-Application Django (AddicTest)

Page d'accueil :



Page About : Qu'est-ce que AddicTest ? : contenant des informations sur les drogues

ADDICTEST[Accueil](#)[Qu'est-ce que AddicTest ?](#)[Prédiction du risque d'addiction](#)[Espace Admin](#)

Qu'est-ce que la AddicTest ?

Bienvenue sur AddicTest

— — NOS SERVICES — —

Prédictions que nous offrons



Consommation d'Alcool

L'alcool est une substance obtenue par fermentation de végétaux (fruits, céréales grains ou racines) riches en sucres comme : Le raisin pour le vin ou le cognac, La pomme (cidre) .



Consommation de Nicotine

La nicotine est un alcaloïde naturellement présent à forte concentration dans les feuilles du tabac. Notons qu'elle est aussi présente dans d'autres plantes comme la tomate et l'aubergine. Et sa toxicité protège la plante des insectes.



Consommation de Cannabis

Le Cannabis est considérée comme un dépressur, un stimulant et un hallucinogène. Le médicament déclenche la libération de dopamine dans le cerveau et produit un "high" relaxant. L'ingrédient actif de l'herbe est le THC, qui produit le high que les gens ressentent lorsqu'ils l'utilisent.



Consommation d'Amphet

Les amphétamines sont des stimulants, ce qui signifie qu'elles accélèrent la vitesse de transmission des messages entre votre cerveau et votre corps. Les stimulants sur ordonnance traitent le ADHD, la narcolepsie et l'obésité.

[→](#)



Consommation de Cocaïne

La cocaïne est une substance hautement addictive et illégale qui entraîne une augmentation des niveaux d'énergie, de la vigilance, de l'euphorie et parfois de l'anxiété. Il est classé comme un stimulant du SNC (système nerveux central) et un anesthésique local.



Consommation d'Ecstasy

L'ecstasy est considérée comme une « drogue de fête » parce que la plupart des consommateurs en consomment pour améliorer leur humeur pendant qu'ils font la fête. Le médicament est fabriqué par l'homme (synthétique) et peut avoir un impact négatif sur votre cerveau, votre corps et vos comportements.

© AddicTest, All Right Reserved.

Accueil

Page de prédiction du risque d'addiction : Saisi d'informations personnels et de tests psychologiques

ADDICTEST

[Accueil](#) [Qu'est-ce que AddicTest ?](#) [Prédiction du risque d'addiction](#)

Espace Admin

Prédiction du risque d'addiction

Simulez votre risque avec notre IA

— — ADDICTEST — —

Simulez votre risque d'addiction

L'outil interactif d'AddicTest vous propose de simuler votre risque d'addiction en utilisant des modèles de **Machine Learning**.

Le modèle utilisé a été réalisé en recueillant des informations personnelles et des résultats de tests psychologiques. Ainsi que des informations sur le niveau de consommation de drogues légales et illégales de plus de 1880 participants, tel que, **l'alcool, benzodiazépine, nicotine, cannabis, cocaïne, crack, ecstasy, héroïne, etc.**

Votre âge

18-24

▼

Votre sexe

Male

▼

Niveau d'éducation

J'ai quitté l'école avant 16 ans

▼

Votre Pays de résidence

Albania

▼

Votre ethnicité

Asiatique

▼

Test Psychologique

Névrosisme

0.00004

Extraversion

0.00305

Ouverture à l'expérience

0.0633

Amabilité

0.12

Conscience

0.704

Impulsivité

0.5338

Recherche de sensations

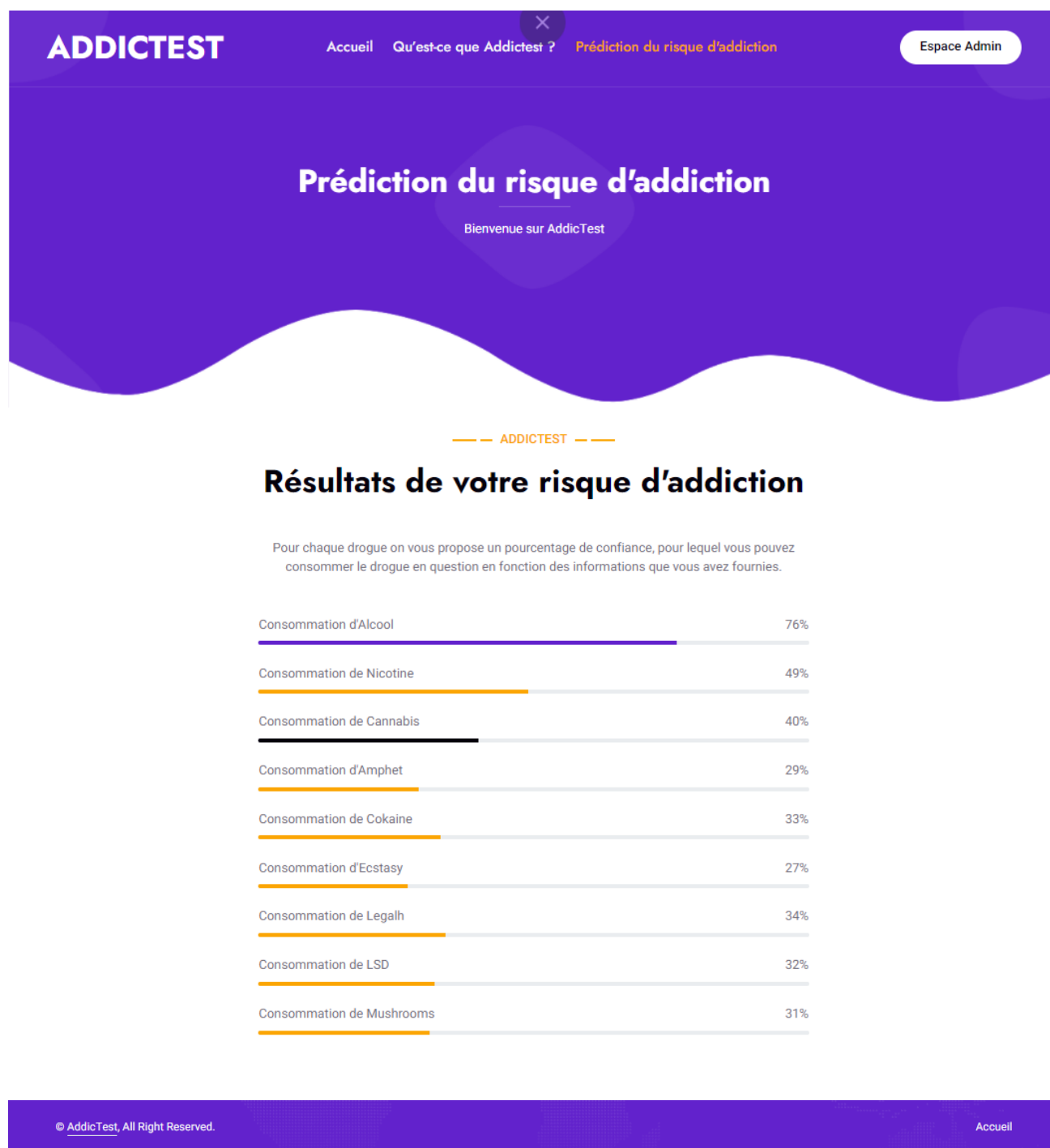
0.2343

Obtenir les résultats

© AddicTest, All Right Reserved.

Accueil

Page des résultats de prédiction du risque d'addiction



Espace Admin : Gestion de résultat de prédiction des utilisateurs

Django administration

WELCOME, ADMIN. [VIEW SITE](#) / [CHANGE PASSWORD](#) / [LOG OUT](#)

Home > Predict > User_experiences

AUTHENTICATION AND AUTHORIZATION

Groups

+ Add

Users

+ Add

PREDICT

User_experiences

+ Add

Select user_experience to change

ADD USER_EXPERIENCE +

Action: Go 0 of 11 selected

☐ USER_EXPERIENCE

☐ User_experience object (11)

☐ User_experience object (10)

☐ User_experience object (9)

☐ User_experience object (8)

☐ User_experience object (7)

☐ User_experience object (6)

☐ User_experience object (5)

☐ User_experience object (4)

☐ User_experience object (3)

☐ User_experience object (2)

☐ User_experience object (1)

11 user_experiences

AUTHENTICATION AND AUTHORIZATION

Groups

+ Add

Users

+ Add

PREDICT

User_experiences

+ Add

Change user_experience

HISTORY

User_experience object (11)

Age: 18-24

Gender: M

Education: J'ai quitté l'école avant 16 ans

Contry: Albania

Ethnicity: 0.0

Nscore: 0.00004

Escore: 0.00305

Oscore: -0.06330

Ascore: 0.12000

Conclusion Générale

Ce projet constitue, pour nous, une grande valeur ajoutée dans notre vie professionnelle. En effet, il nous a permis :

- D'adapter les acquis théoriques et pratiques que nous avons appris durant notre formation d'ingénieur (l'apprentissage automatique et génie logiciel et Python)
- Effectuer une mise en œuvre réelle d'un projet intéressant.
- De vivre l'expérience d'un vrai travail en groupe.
- De découvrir les nouvelles technologies Web et de Data science en général qui font l'actualité dans le domaine d'intelligence artificielle.
- D'élargir notre esprit et notre réflexion en tant que future ingénieur en data numérique et intelligence artificielle.

En bref, c'était une expérience enrichissante sur tous les points qu'ils soient théoriques ou pratiques. Cette expérience nous a été très profitable au niveau de l'amélioration de nos connaissances.

Bibliographie

“Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow - Concepts, Tools, and Techniques to Build Intelligent System (2019) “ *Aurélien Géron*.

Webgraphie

<https://www.kaggle.com/>

<https://www.djangoproject.com/start/>

<https://www.atlassian.com/fr/agile/scrum>

<https://scikit-learn.org/stable/>