



## Rapport du Mini Projet Sujet : Application de machine learning

Sous l'encadrement de Pr. Sanae Khali Issa

Préparé par  
El ghazi Loubna  
Zaoui Hanane

# SOMMAIRE

<b>1.</b>	<b>INTRODUCTION AU SUJET .....</b>	<b>3</b>
1.1.	Introduction.....	3
1.2.	But de ce projet.....	3
<b>2.</b>	<b>MODÉLISATION.....</b>	<b>4</b>
2.1.	Diagramme de Use Cases.....	4
2.2.	Diagramme de séquences :Création du dataset : .....	5
2.3.	Diagramme de séquences : Choix du modèle.....	6
2.4.	Diagramme d'activité.....	7
<b>3.</b>	<b>REALISATION .....</b>	<b>7</b>
3.1.	Outils Utilisés .....	7
3.2.	Structure de l'application .....	9
3.3.	Fonctionnalités de l'application .....	9
<b>4.</b>	<b>CONCLUSION.....</b>	<b>15</b>
<b>5.</b>	<b>WEBOGRAPHIE.....</b>	<b>16</b>

# 1. INTRODUCTION AU SUJET

## 1.1. Introduction

Le projet que nous présentons s'inscrit dans le cadre de fournir une expérience enrichissante aux étudiants dans les domaines de la programmation et de l'apprentissage automatique. L'objectif principal de ce projet est de développer une application informatique polyvalente en utilisant le langage de programmation Python. Cette application sera conçue pour analyser des ensembles de données, qu'ils soient qualitatifs ou quantitatifs, en mettant en œuvre divers algorithmes de machine learning.

L'apprentissage automatique, également connu sous le nom de machine learning, est un domaine en pleine expansion qui offre des opportunités passionnantes pour exploiter les données et générer des modèles prédictifs. Dans ce contexte, notre application se concentrera sur plusieurs algorithmes de machine learning tels que la régression linéaire sous ses différentes formes (simple, polynomiale, multiple, logistique), les arbres de décision, le naïf bayésien, le support vector machine (SVM), Random Forest, et le plus proche voisin.

Cette initiative vise à doter les étudiants des compétences pratiques nécessaires pour aborder des problèmes du monde réel en utilisant des techniques avancées d'analyse de données. En mettant l'accent sur la programmation en Python, un langage de plus en plus prisé dans le domaine de la science des données, nous visons à fournir une expérience pratique et pertinente qui complète efficacement les connaissances théoriques acquises au cours de leurs études.

Ce rapport détaillera le processus de conception, de développement et d'implémentation de l'application, en mettant en lumière les choix de conception, les défis rencontrés et les solutions apportées. De plus, nous examinerons en détail chaque algorithme de machine learning intégré dans l'application, en expliquant leur fonctionnement, leur utilité, et en illustrant comment ils sont mis en œuvre pour atteindre des résultats significatifs dans l'analyse de données.

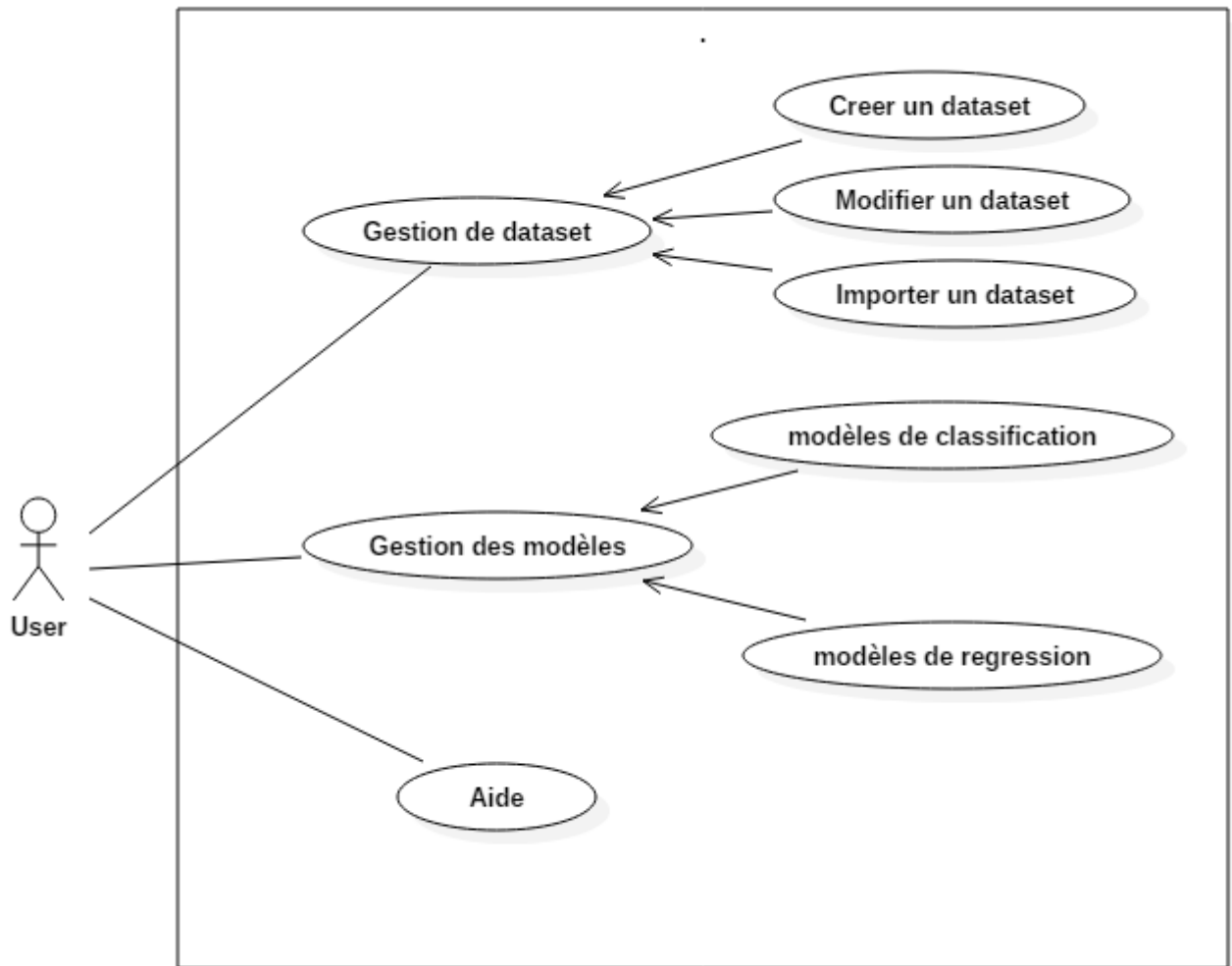
## 1.2. But de ce projet

Le projet à réaliser vise à fournir aux étudiants une expérience pratique dans le domaine de la programmation et de l'apprentissage automatique en développant une application informatique polyvalente utilisant le langage de programmation Python.

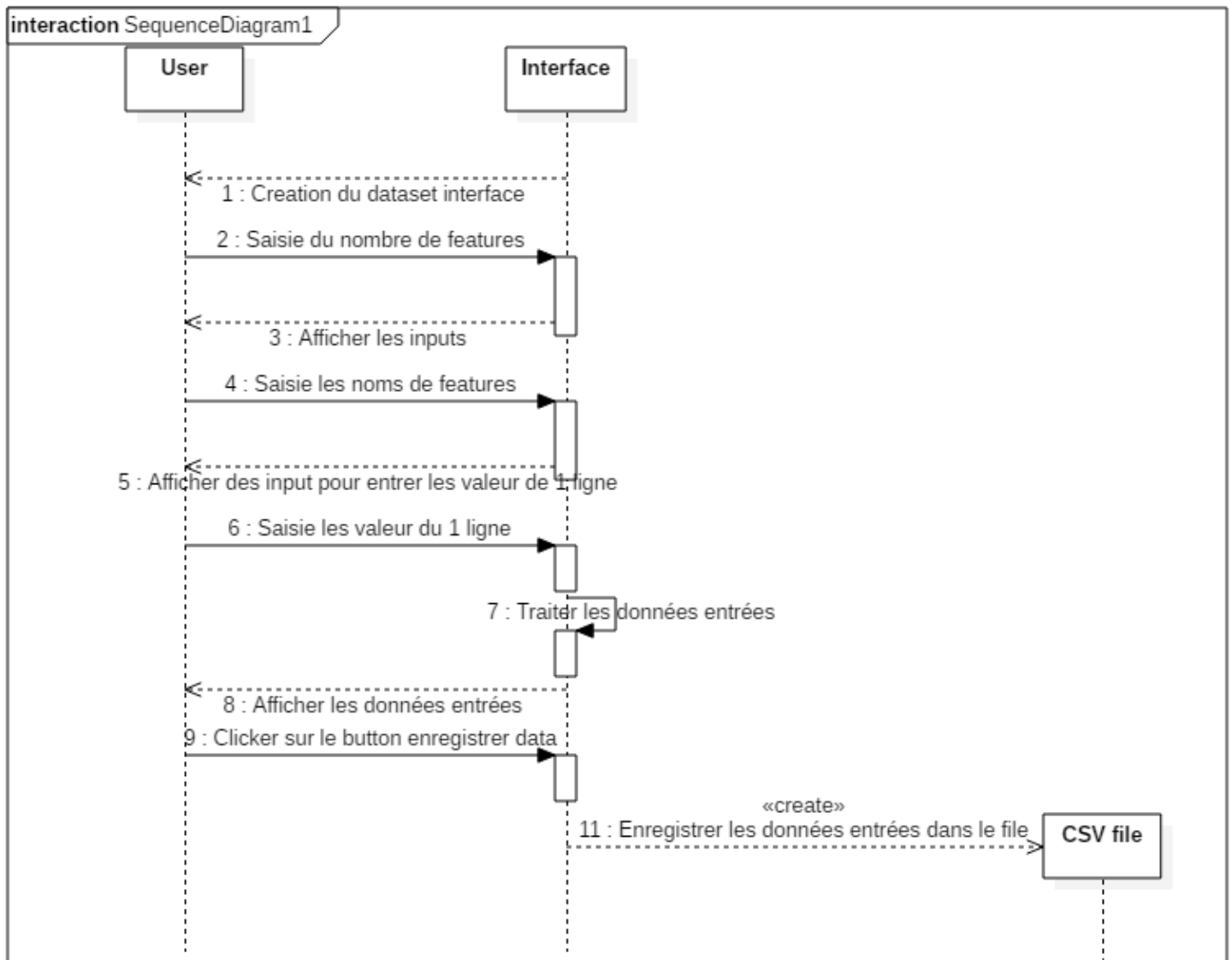
Cette application consiste à analyser un ensemble de données qualitatives ou quantitatives en implémentant certains algorithmes de Machine Learning : la régression linéaire avec ses différentes catégories (simple, polynomiale, multiple, logistique), les arbres de décision, le naïf bayésien, le support vector machine (SVM), Random forest, le plus proche voisin et Kmeans.

## 2. MODÉLISATION

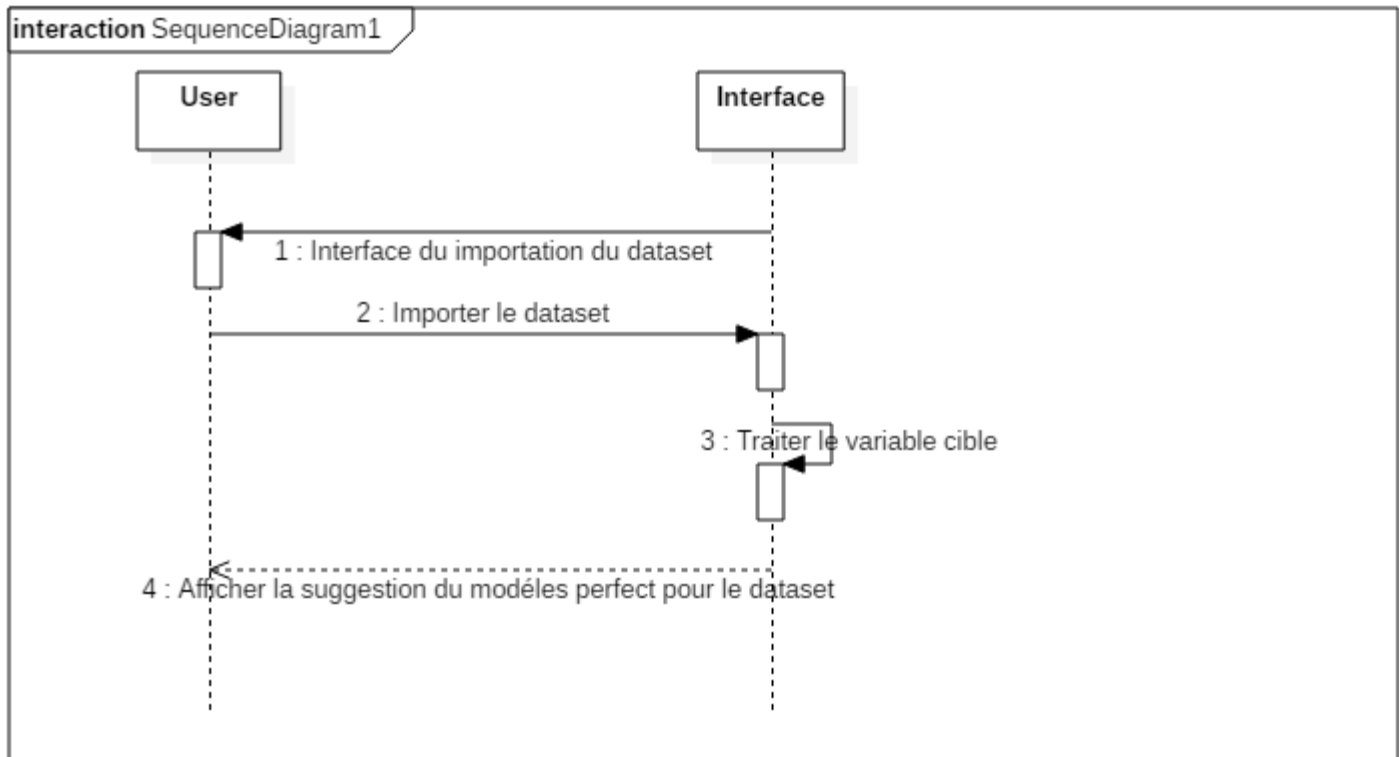
### 2.1. Diagramme de Use Cases



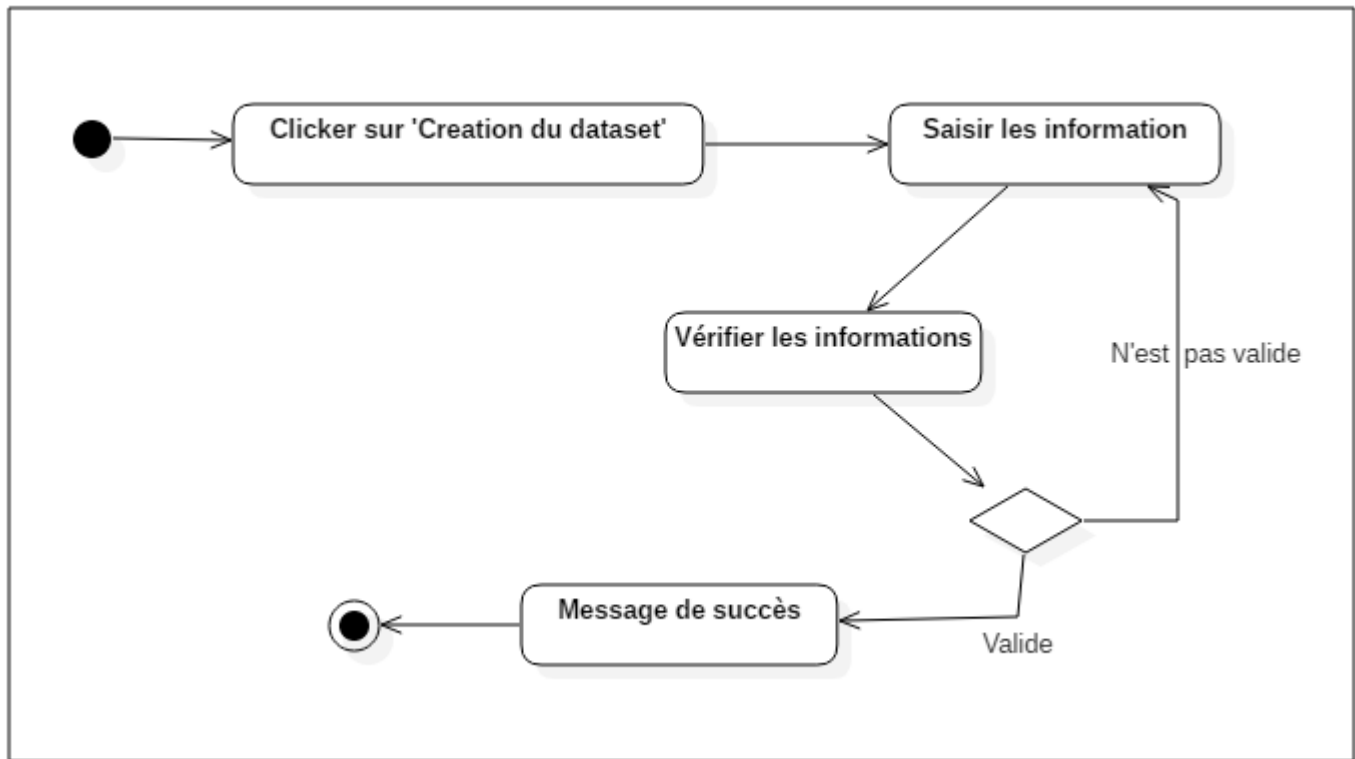
## 2.2. Diagramme de séquences :Création du dataset :



## 2.3. Diagramme de séquences : Choix du modèle



## 2.4. Diagramme d'activité



## 3. REALISATION

### 3.1. Outils Utilisés

Pour programmer notre application, on a utilisé la programmation procédurale en utilisant Python, et en intégrant des modules et fonctions pour créer ses différentes pages.

Dans notre application, on a utilisé les outils suivants :



Python est un langage de programmation interprété, polyvalent et facile à lire, conçu pour favoriser une syntaxe claire et un code lisible. Il offre une approche conviviale pour le développement d'applications, de scripts et d'automatisations, ce qui en fait un choix populaire tant pour les débutants que pour les développeurs expérimentés.

De plus, on a utilisé les modules et bibliothèques de PYTHON suivantes :

**io (module dans la bibliothèque standard) :**

**Description :** Le module io fait partie de la bibliothèque standard de Python et fournit des outils pour gérer les opérations d'entrée/sortie (I/O), y compris la lecture et l'écriture de fichiers.

**Joblib :**

**Description :** Joblib est une bibliothèque en Python qui fournit des fonctionnalités pour la parallélisation et la distribution de tâches, en particulier pour les tâches qui impliquent des calculs intensifs.



Streamlit est une bibliothèque open-source en Python qui permet de créer des applications web de manière simple et rapide pour l'analyse de données et la visualisation.



NumPy est une bibliothèque fondamentale pour le calcul numérique en Python. Elle offre des structures de données, des fonctions mathématiques et des outils pour travailler avec des tableaux multidimensionnels et des matrices.



Matplotlib est une bibliothèque de visualisation 2D en Python. Elle permet de créer des graphiques, des diagrammes, des histogrammes, etc., de manière flexible.



Seaborn est une bibliothèque de visualisation de données basée sur Matplotlib. Elle fournit une interface haut niveau pour créer des graphiques statistiques informatifs et attrayants.



Pandas est une bibliothèque qui offre des structures de données performantes et faciles à utiliser, notamment les DataFrame. Les DataFrames sont similaires à des tableaux ou des feuilles de calcul, et Pandas offre des outils pour manipuler, analyser et nettoyer des données.





"scikit-learn", souvent abrégé en "sklearn", est une bibliothèque open-source en Python qui offre des outils simples et efficaces pour l'apprentissage automatique (machine learning). Elle est construite sur les bibliothèques NumPy, SciPy et Matplotlib, et propose des modules pour diverses tâches liées à l'apprentissage automatique, y compris la classification, la régression, le clustering, la réduction de dimensionnalité, la sélection de modèle, et plus

## 3.2. Structure de l'application

La fonction principale de notre application est : **Home.py** : pour l'exécuter, vous devez taper

La commande : **streamlit run App.py**.

Ce fichier contient le menu principal de l'application, et importe toutes les pages de l'app qui sont importés dedans.

Les pages de l'application sont représentées dans un dossier nommé **compo** comme suit :

**Aide.py, Regression.py, Modify\_data.py, pageIntro.py, DataEntry.py, Classification.py, Clustering.py**

Chaque page représente une fonction qui contient leurs fonctionnalités et importés comme Modules.

Pour le Design de l'application, on a utilisé en plus des fonctionnalités de **Streamlit**, du CSS (**Cascading Style Sheets**).

## 3.3. Fonctionnalités de l'application

### MENU DE NAVIGATION :

#### 1. Accueil



La page d'accueil offre un ensemble de fonctionnalités pour préparer efficacement les données en vue de l'analyse. Les actions disponibles comprennent :

**Création du dataset** : Permet à l'utilisateur de générer un ensemble de données personnalisé adapté à ses besoins d'analyse.

L'ensemble de données crée s'enregistre sous forme d'un fichier csv après avoir choisi son nom ,il peut donc l'utiliser pour les modèles .

**Modification des données** : Offre la possibilité de manipuler et ajuster les données existantes(importées) en fonction des exigences spécifiques de l'utilisateur.

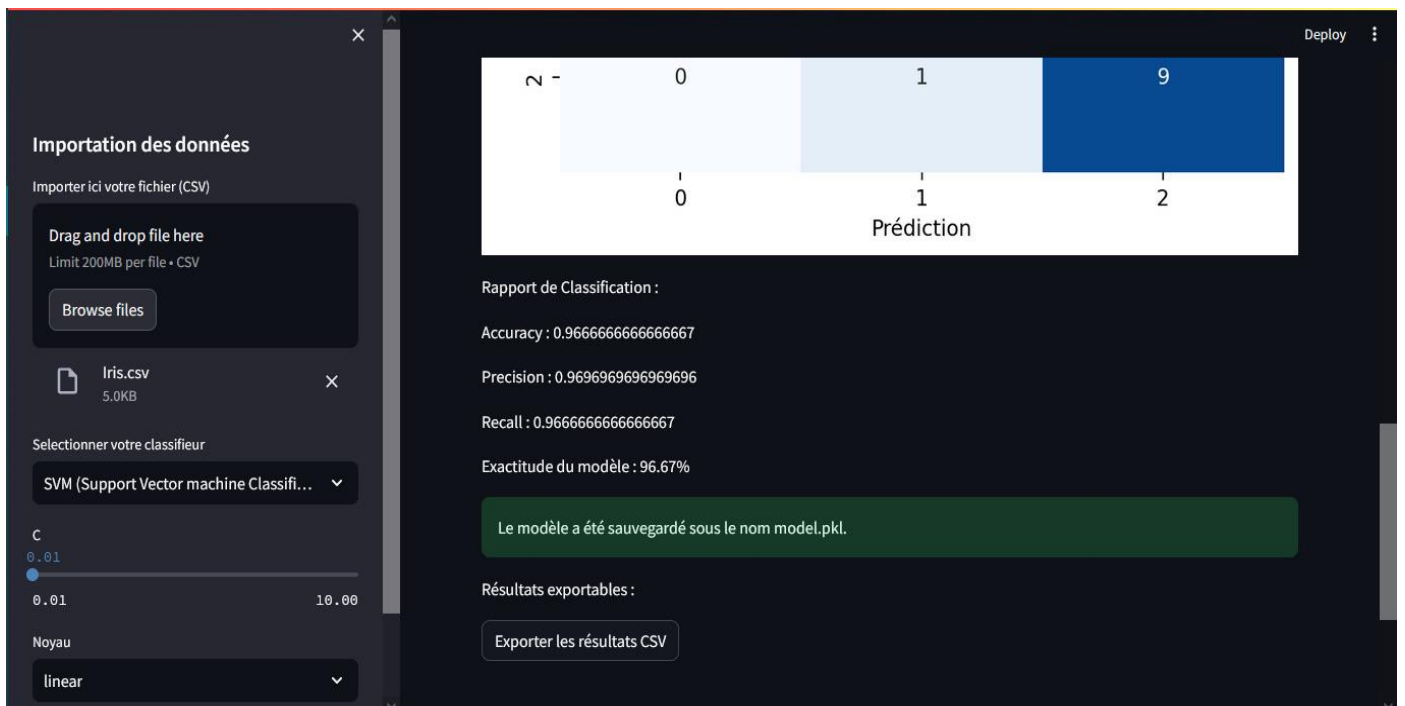
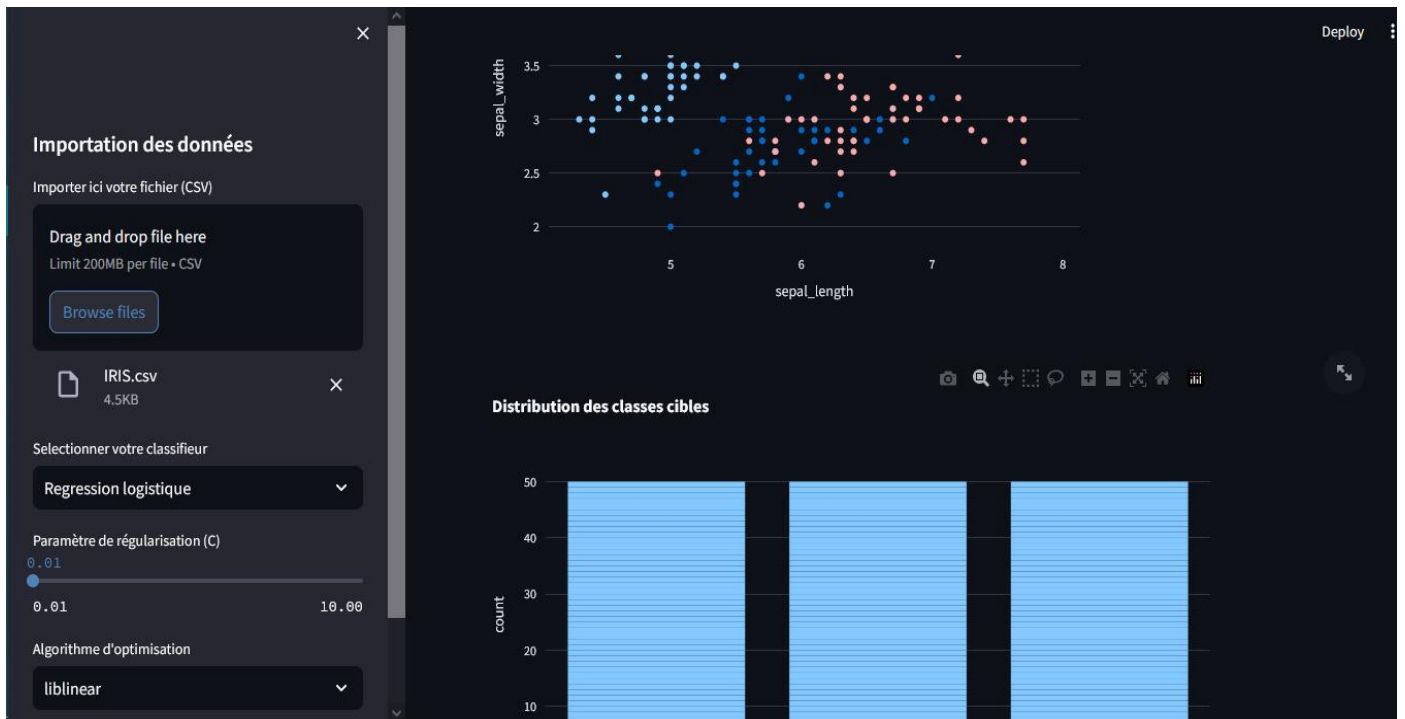
**Importation des données** : Permet d'importer facilement des ensembles de données pré-existants. Une fois les données importées, l'application guide l'utilisateur vers le modèle approprié (Classification ou Régression) ou (Clustering) en fonction des caractéristiques des données (En calculant et comparant les variances des variables dans les fonctions des modèles).

## 2. Modèles

The screenshot shows the 'Modèles' section of the application. On the left, there is a sidebar with 'Importation des données' (Data Import) and 'Selectionner votre classifieur' (Select your classifier). The main area displays a welcome message 'Bienvenue dans votre application de Machine Learning !' and a navigation bar with 'Accueil', 'Modèles', and 'Aide'. Below the navigation bar, there is a dropdown menu for 'Classification' and a list of classifiers: 'Classification', 'Régression', and 'Clustering'. The 'Visualisation de l'ensemble de données' (Data Set Visualization) section is highlighted, showing a message: 'Veuillez charger ou générer des données pour continuer.' (Please load or generate data to continue).

The screenshot shows the 'Modèles de classification' section of the application. The sidebar on the left lists various classifiers: 'Regression logistique', 'Naive Bayes', 'SVM (Support Vector machine Classifier)', 'Arbres de décisions', 'KNN (le plus prochain voisin)', 'Random Forest', and 'SVM (Support Vector machine Classifi...)'. The 'C' parameter is set to 0.01, and the 'Noyau' (Kernel) is set to 'linear'. The main area displays the same welcome message and navigation bar. Below the navigation bar, there is a dropdown menu for 'Classification' and a list of classifiers: 'Classification', 'Régression', and 'Clustering'. The 'Visualisation de l'ensemble de données' (Data Set Visualization) section is highlighted, showing a table of data points.

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa



The screenshot shows a web-based machine learning interface. On the left, under 'Importation des données', there is a file upload section with a 'Browse files' button and a file named 'IRIS.csv' (4.5KB) is shown. Below this, a classifier is selected as 'SVM (Support Vector machine Classifi...)' and a kernel is set to 'linear'. On the right, under 'Sélectionnez les fonctionnalités à utiliser', four features are selected: 'sepal\_length', 'sepal\_width', 'petal\_length', and 'petal\_width'. Below these, input fields are provided for each feature: 'sepal\_length' (2), 'sepal\_width' (34), 'petal\_length' (5), and 'petal\_width' (2). A 'Tester' button is located below the inputs. At the bottom right, the prediction result is shown as 'Iris-setosa'.

La section "Modèles" est divisée en trois parties pour fournir une analyse complète :

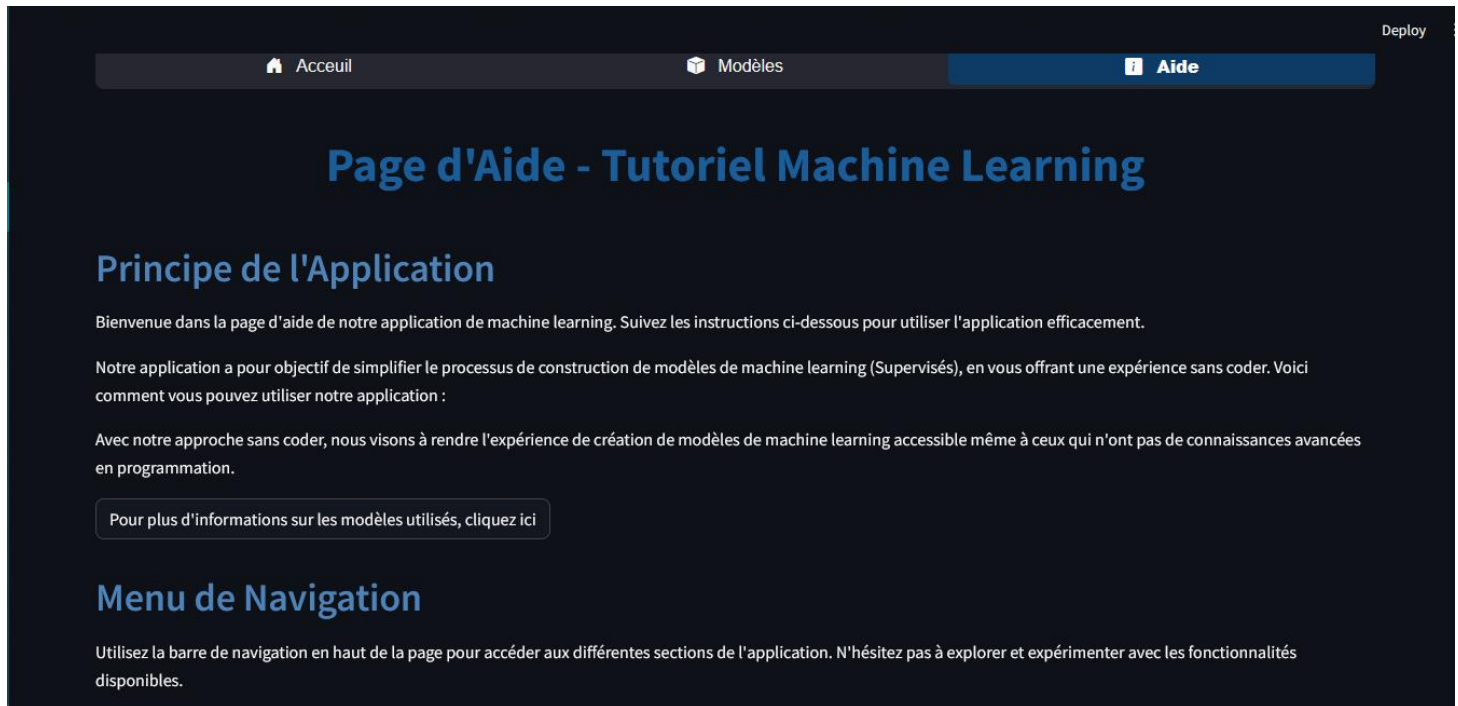
**Visualisation de l'ensemble de données :** Permet à l'utilisateur de visualiser divers graphiques pour mieux comprendre la distribution et les relations dans les données.

Le nettoyage des données, la préparation et la transformation de données, et la gestion des valeurs manquantes et aberrantes sont tous présents dans cette partie.

**Résultats du modèle :** Après la sélection d'un modèle et l'ajustement des paramètres (pour chaque modèle), cette section présente la performance du modèle ainsi que diverses métriques. De plus, l'utilisateur peut exporter les résultats à tout moment pour une référence ultérieure.

**Test :** Offre la possibilité d'effectuer des tests supplémentaires pour évaluer la robustesse du modèle et affiner les paramètres si nécessaire.

### 3. Aide



La section "Aide" est dédiée à fournir des ressources d'assistance aux utilisateurs. Vous y trouverez :

**Documentation :** Des guides détaillés et des tutoriels pour vous aider à utiliser efficacement chaque fonctionnalité de l'application avec une section de **Comparaison des Modèles Supervisés et Non Supervisés** pour fournir une description et une comparaison des modèles supervisés et non supervisés. Cela permet aux utilisateurs de comprendre les différences entre ces deux approches et de choisir la méthode la plus appropriée en fonction de leurs besoins spécifiques en analyse de données.

## 4. CONCLUSION

En conclusion, notre application d'analyse de données représente un outil complet et convivial pour les utilisateurs cherchant à explorer, analyser et modéliser leurs ensembles de données. De la préparation initiale des données à l'évaluation des modèles, notre plateforme offre une expérience intégrée visant à simplifier les processus complexes de l'analyse de données.

La barre de navigation intuitive permet une navigation fluide entre les différentes sections, offrant ainsi un accès rapide aux fonctionnalités clés. La section "Aide" a été spécialement conçue pour assister les utilisateurs à chaque étape en offrant des guides détaillés.

L'ajout de la section comparative entre les modèles supervisés et non supervisés dans l'aide vise à éclairer les utilisateurs sur les différentes approches disponibles, les aidant ainsi à choisir la méthode la plus adaptée à leurs besoins spécifiques.

Nous aspirons à fournir une solution qui non seulement simplifie l'analyse de données, mais également encourage l'exploration et l'apprentissage continu.

## 5. WEBOGRAPHIE

Streamlit documentation : <https://docs.streamlit.io/>

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

Q&A platform : <https://stackoverflow.com/>

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

Matplotlib 3.8.2 documentation : <https://matplotlib.org/stable/index.html>

Streamlit discord : <https://discuss.streamlit.io/t/css-styling/35243>