

Soutenance Projet 3A

Création d'une plateforme d'analyse de données
et de machine learning



Etudiant : Paul Aldeano
Tutrice : Dominique Benmouffek

Sommaire

- Historique du projet et objectifs
- Travail sur la plateforme de M. Couillard
- Réalisation d'une deuxième plateforme
- Perspectives d'amélioration

Historique du projet

Datanalyze

Datasets

Tables

Analyze

Machine Learning

Datasets index

Boston Housing

This dataset contains information collected by the U.S Census Service concerning housing in the area of Boston Mass. It was obtained from the StatLib archive (<http://lib.stat.cmu.edu/datasets/boston>), and has been used extensively throughout the literat...

EditDelete

Wine

These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. I think that t...

EditDelete

Heart Disease Dataset

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence o...

EditDelete

Vehicule Dataset

This data was originally gathered at the TI in 1986-87 by JP Siebert. It was partially financed by Barr and Stroud Ltd. The original purpose was to find a method of distinguishing 3D objects within a 2D image by application of an ensemble of shape featu...

EditDelete

CO2 PPM

CO2 PPM - Trends in Atmospheric Carbon Dioxide. Data are sourced from the US Government's Earth System Research Laboratory, Global Monitoring Division. Two main series are provided: the Mauna Loa series (which has the longest continuous series since 195...

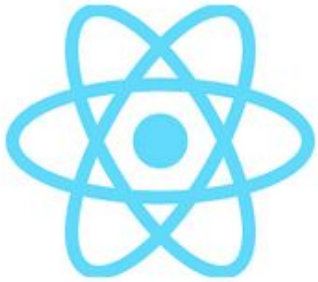
EditDelete

+

Objectifs principaux

- Améliorer l'ergonomie de la plateforme
- Étendre ses capacités à des jeux de données plus ambitieux

Travail sur la plateforme de M. Couillard



ReactJS
Interface utilisateur



Python
Backend



Elasticsearch
Base de données



Docker
Déploiement

Travail sur la plateforme de M. Couillard

Dataset Explorer

Select a dataset

Test

<!DOCTYPE html>

<html lang="en" data-color-mode="auto" data-light-theme="light" data-dark-theme="dark">

<head>

<meta charset="utf-8">

<link rel="dns-prefetch" href="https://github.githubassets.com">

<link rel="dns-prefetch" href="https://avatars.githubusercontent.com">

<link rel="dns-prefetch" href="https://github-cloud.s3.amazonaws.com">

<link rel="dns-prefetch" href="https://user-images.githubusercontent.com/">

<link rel="preconnect" href="https://github.githubassets.com" crossorigin>

<link rel="preconnect" href="https://avatars.githubusercontent.com">

<link crossorigin="anonymous" media="all" integrity="sha512-dkuYFW+ra8yYSt342e5pJEsIPsJMcrMvNxiVZMyM/X+W/JHDPvoCuGq3LFojl7B0dQWwZNRiPMnbi9IfUgTaA==" rel="stylesheet" href="https://github.githubassets.com/assets/light-764b98156fab6b6cc984addf8d9ee6924.css" /><link crossorigin="anonymous" media="all" integrity="sha512-UrAu23+eyncWwaQFwsLbgSKtrmLb2aH1bcT4hJnnRdkaPuY1eu9bumt33FyHHFDX8skTUNWNklsMCz7FWQQHwA==" rel="stylesheet" href="https://github.githubassets.com/assets/dark-52b02edb7f9eca7716bda405c2c2db81.css" /><link data-color-theme="dark_dimmed" crossorigin="anonymous" media="all" integrity="sha512-kyu73YWTu8Fu2e7p+Hv094CRhaTvr8yy95vc1SQ2+MeWVWwakGelH/lv9yIfaYAb8J3oM6uBLGcn1kS6M1GxBcq==" rel="stylesheet" data-href="https://github.githubassets.com/assets/dark_dimmed-932bbdd85ad53c16ed9eee9f87bf4f7.css" /><link data-color-theme="dark_high_contrast" crossorigin="anonymous" media="all" integrity="sha512-jZSKF7Gx8T/AFthOOCUKWWpG5EBIIZb+tiYu8KgP/kizn7fpXEiXjcB73GTZ69wSVV5Z6Y1Cw286qP7pVZr0gg==" rel="stylesheet" data-href="https://github.githubassets.com/assets/dark_high_contrast-8d948a17b1b1f13fc016d84ed0252459.css" /><link data-color-theme="dark_colorblind" crossorigin="anonymous" media="all" integrity="sha512-E02WD8opZpPy6Lm9diUSIHQgXtLmzi1KxMnaN/SA7k6iLsvpNjjpKBPU1sC98MitAOcCNle6ozqY8+pHnrHZg==" rel="stylesheet" data-href="https://github.githubassets.com/assets/dark_colorblind-134d960fca2964fa58bba2ccf5d95448.css" /><link data-color-theme="light_colorblind" crossorigin="anonymous" media="all" integrity="sha512-VWdBPHZj3WCdwaO0N2W8yvDZt7TNZohRIYK4sj5U56485CWazxnLr4p3DUBeqn2+eSj3CYPw4+DzmwH0wew==" rel="stylesheet" data-href="https://github.githubassets.com/assets/light_colorblind-5567413c7663dd6083c1a3b43765bcca.css" /><link data-color-theme="light_high_contrast" crossorigin="anonymous" media="all" integrity="sha512-

Travail sur la plateforme de M. Couillard



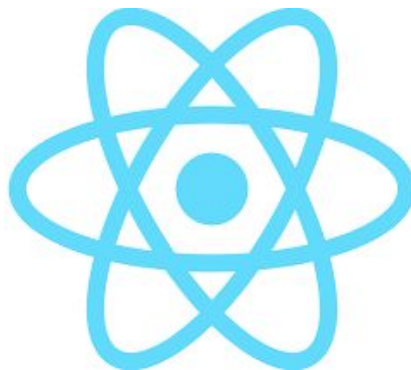
Environnement virtuel



Dataset

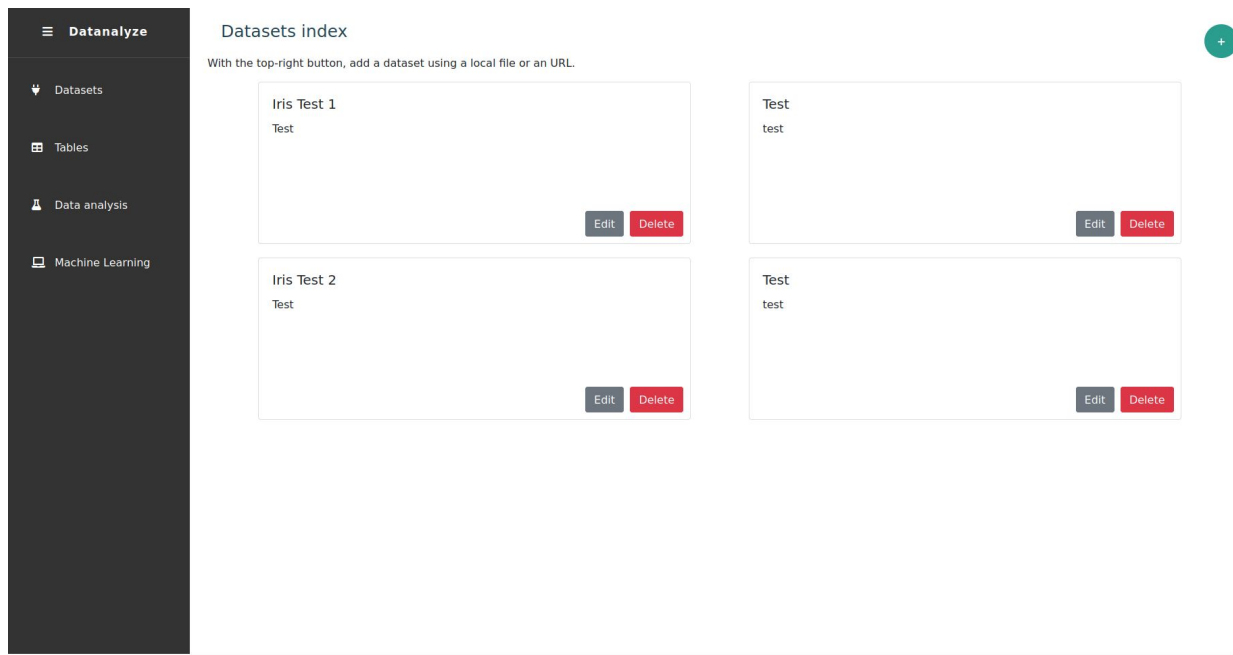
Machine locale

Travail sur la plateforme de M. Couillard

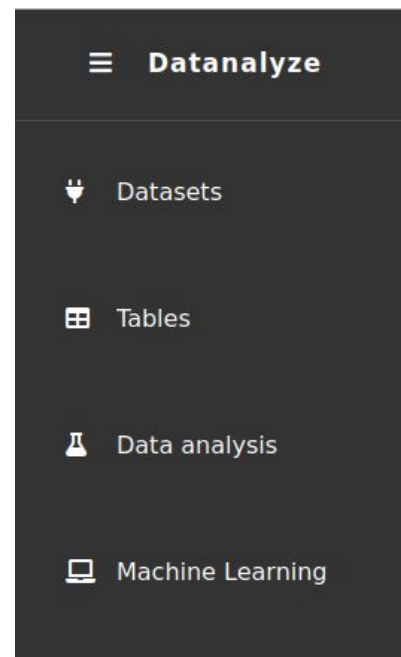


CDBreact

Travail sur la plateforme de M. Couillard



The screenshot shows the Datanalyze platform interface. On the left is a dark sidebar with a menu containing: **Datanalyze**, **Datasets** (with a plug icon), **Tables** (with a grid icon), **Data analysis** (with a flask icon), and **Machine Learning** (with a laptop icon). The main content area is titled "Datasets index" and includes a sub-header: "With the top-right button, add a dataset using a local file or an URL." There is a green circular button with a "+" sign in the top right corner. Below the header, there are four dataset cards arranged in a 2x2 grid. The top-left card is titled "Iris Test 1" and contains the text "Test"; it has "Edit" and "Delete" buttons at the bottom right. The top-right card is titled "Test" and contains the text "test"; it also has "Edit" and "Delete" buttons at the bottom right. The bottom-left card is titled "Iris Test 2" and contains the text "Test"; it has "Edit" and "Delete" buttons at the bottom right. The bottom-right card is titled "Test" and contains the text "test"; it also has "Edit" and "Delete" buttons at the bottom right.



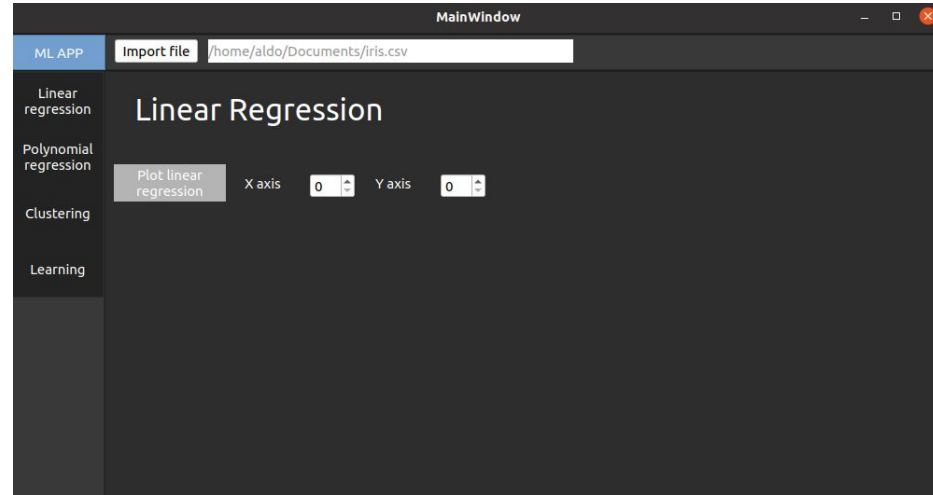
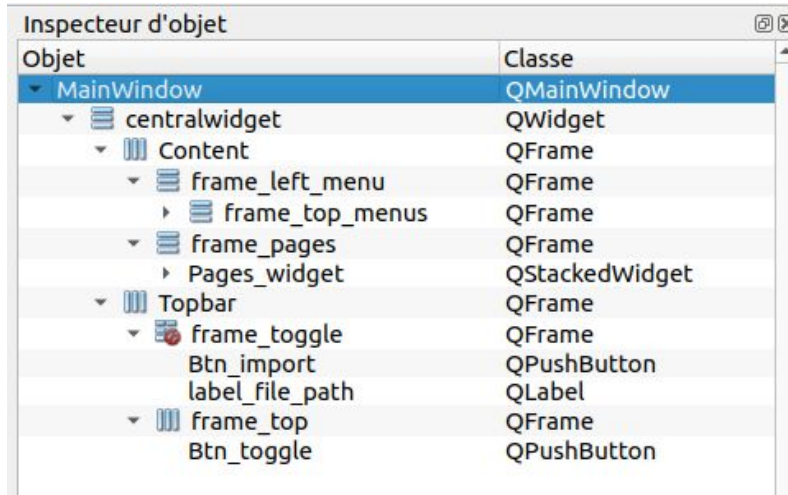
This screenshot shows the sidebar menu of the Datanalyze platform. It features a dark background with the title **Datanalyze** at the top. Below the title are four menu items, each with an icon and text: **Datasets** (plug icon), **Tables** (grid icon), **Data analysis** (flask icon), and **Machine Learning** (laptop icon).

Réalisation d'une deuxième plateforme



Pourquoi python?

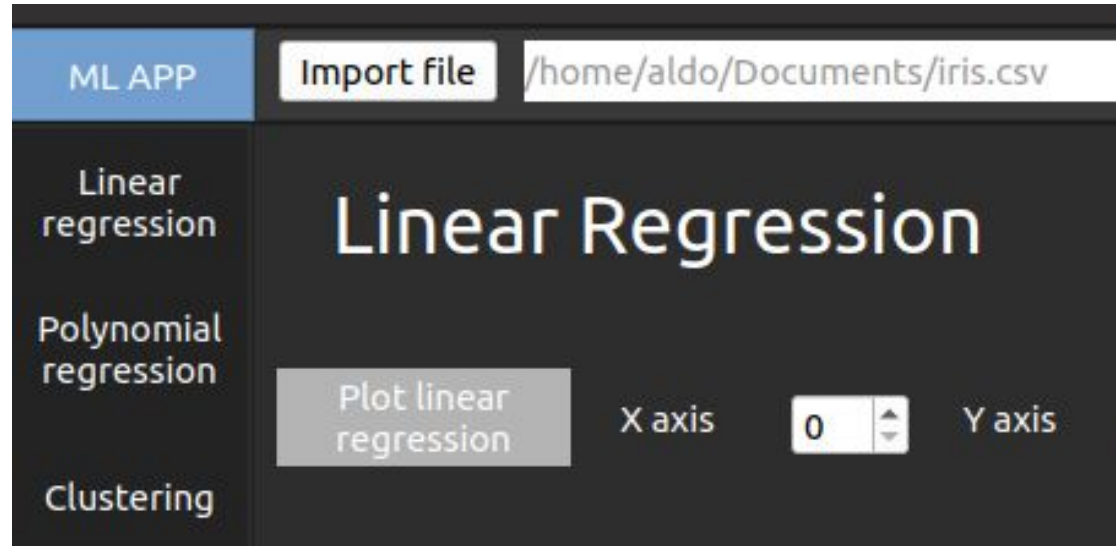
Réalisation d'une deuxième plateforme



Réalisation d'une deuxième plateforme

Fonctionnalités :

- Linear Regression
- Polynomial Regression
- Clustering



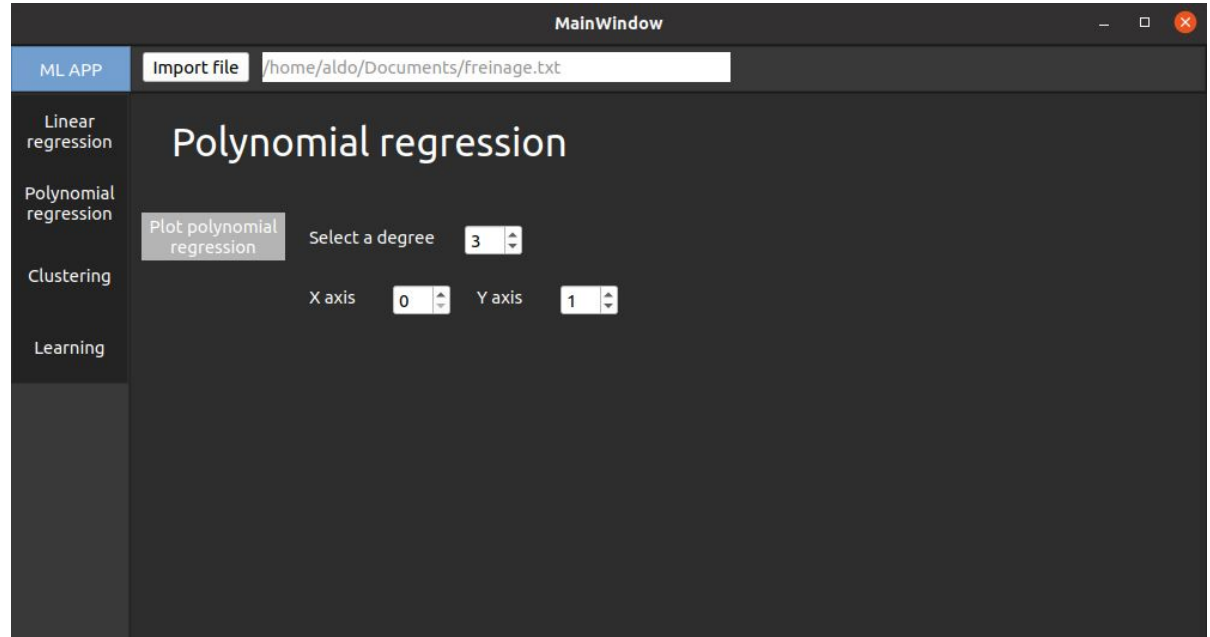
Réalisation d'une deuxième plateforme

Linear regression



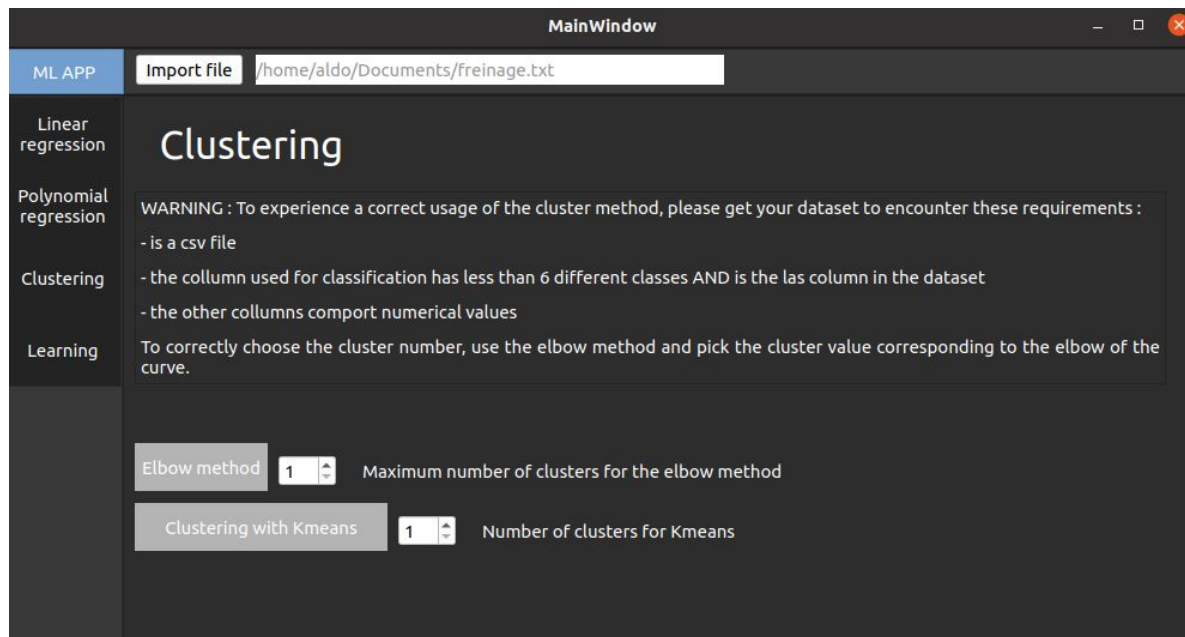
Réalisation d'une deuxième plateforme

Polynomial regression



Réalisation d'une deuxième plateforme

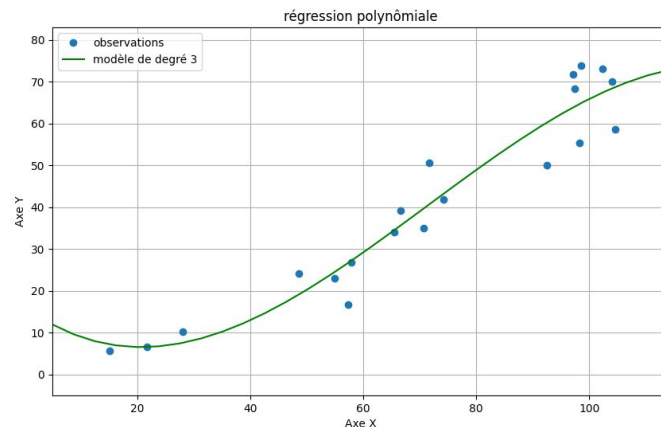
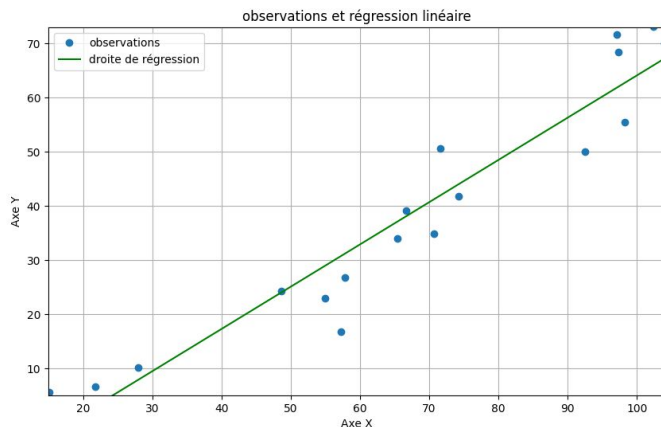
Clustering



Réalisation d'une deuxième plateforme

Résultats obtenus :

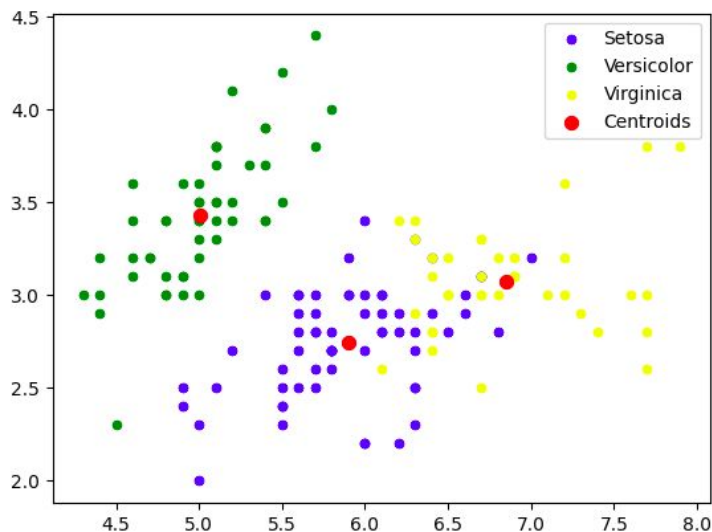
- dataset freinage



Réalisation d'une deuxième plateforme

Résultats obtenus :

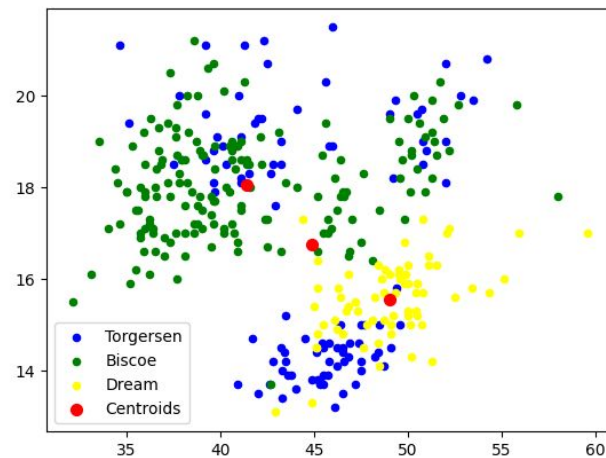
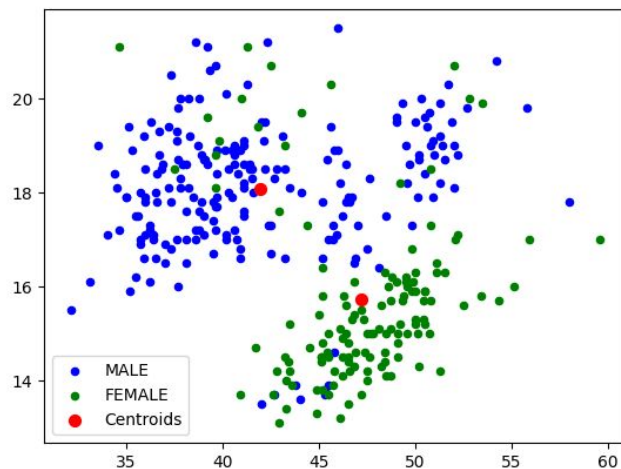
- dataset iris



Réalisation d'une deuxième plateforme

Résultats obtenus :

- dataset pingouins



Perspectives d'améliorations

- Réseau de neurones
- Pré processing des Datasets
- Interface de modification des données

Merci!
