



Introduction :

Ce rapport détaille le développement d'une application informatique utilisant le langage de programmation Python avec l'interface graphique Tkinter. L'objectif principal est de fournir aux étudiants une expérience pratique dans le domaine de la programmation et de l'apprentissage automatique. L'application permet l'analyse d'ensembles de données qualitatives ou quantitatives en mettant en œuvre plusieurs algorithmes de Machine Learning, tels que les arbres de décision, le naïf bayésien, Random forest, le plus proche voisin.

Fonctionnalités Clés :

Interface Utilisateur Intuitive : Création d'une interface conviviale avec Tkinter pour une manipulation efficace des données.

Représentation des Données : Interface graphique pour une représentation claire et explicite des données.

Gestion des Données : Intégration de fonctionnalités pour la préparation et la transformation des données, incluant le nettoyage, la gestion des valeurs manquantes et la normalisation des caractéristiques.

Algorithmes de Machine Learning : Implémentation d'algorithmes variés pour permettre aux utilisateurs de choisir le modèle adapté à leurs données.

Validation des Modèles : Intégration de méthodes d'évaluation pour assurer des performances robustes et fiables.

Visualisation des Résultats : Outils de visualisation tels que graphiques et tableaux pour une interprétation claire des performances des algorithmes.

Exportation des Résultats : Possibilité d'exporter les résultats des analyses, les modèles entraînés et les visualisations.

Documentation et Tutoriels : Encouragement à fournir une documentation complète du code, ainsi que des tutoriels expliquant le fonctionnement de chaque algorithme implémenté dans l'application Tkinter.

Cette section présente de manière visuelle et commentée les solutions envisagées pour répondre aux besoins spécifiques du projet. À travers des captures d'écran de notre application développée avec Tkinter, nous mettons en lumière les différentes fonctionnalités clés et l'interface utilisateur intuitive que nous avons intégré.

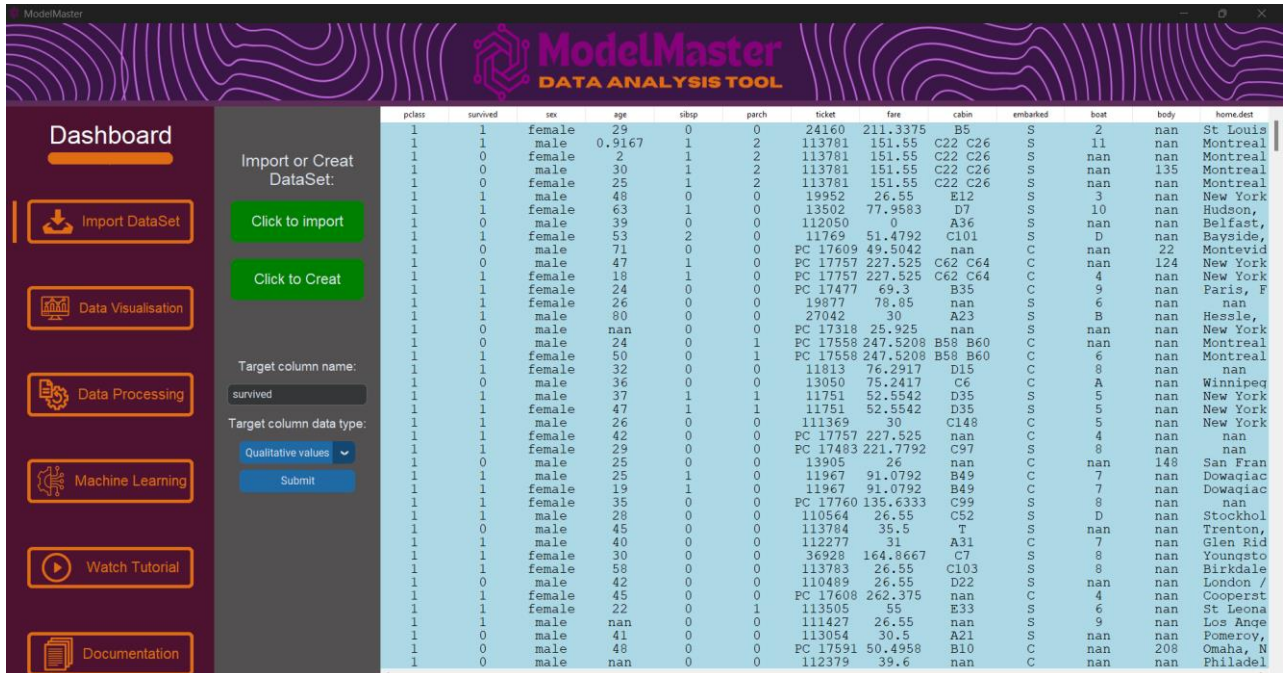
Interface d'Accueil :



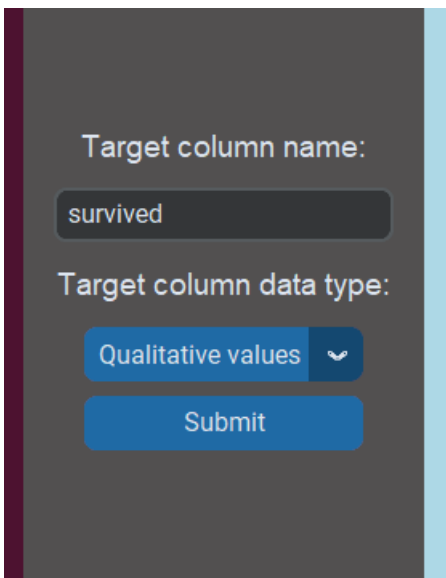
L'interface d'accueil se présente comme un tableau de bord central, offrant une vue d'ensemble de notre application. Elle affiche de manière élégante le nom de l'application et l'équipe de développement, fournissant ainsi une identification immédiate. Au gauche de cette interface, les utilisateurs trouveront des boutons d'accès rapide, facilitant ainsi la navigation vers des sections spécifiques de l'application.

Cette première capture d'écran propose une mise en page épurée avec des boutons clairement libellés, invitant les utilisateurs à explorer les fonctionnalités de l'application de manière intuitive. Ces boutons, soigneusement agencés, dirigent vers des interfaces dédiées pour l'importation et création de jeux de données, la visualisation des données, le traitement des données, l'apprentissage automatique, et enfin, une section spécifique pour la documentation. L'objectif de cette interface d'accueil est de simplifier la navigation et de fournir un point de départ convivial pour les utilisateurs, les guidant efficacement vers les différentes fonctionnalités de l'application.

Importation de Données :



L'interface d'Importation de Données offre une expérience conviviale pour les utilisateurs souhaitant intégrer des datasets dans l'application. De manière intuitive, cette interface propose deux boutons distincts : l'un dédié à l'importation des datasets depuis l'ordinateur de l'utilisateur, il affiche instantanément les lignes et colonnes du dataset après d'importation, et l'autre bouton pour créer un nouveau Dataset.

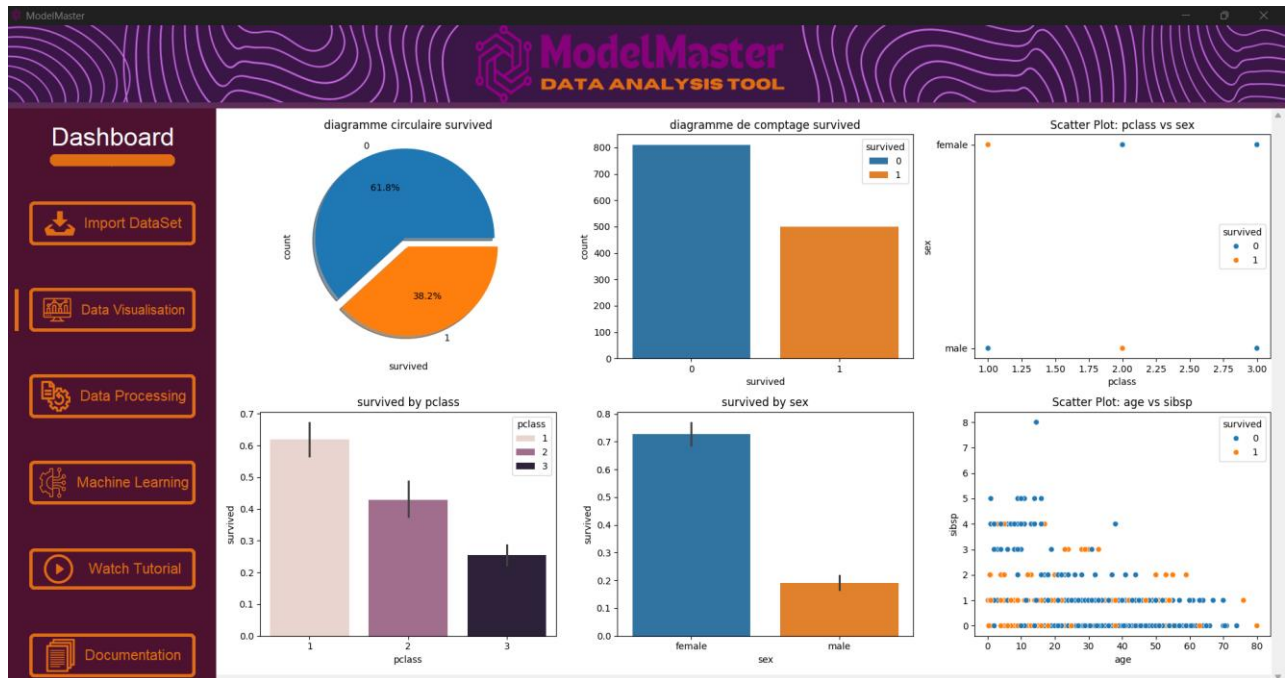


Une fonctionnalité particulièrement utile est la présence d'une case permettant de saisir le nom de la colonne cible (Target), avec la possibilité de déterminer le type de ce dernier (qualitatives ou quantitatives) pour déterminer le type de problème (Régression ou Classification) (si l'utilisateur n'est pas défini le colonne cible ou n'est pas déterminer le type de ce dernier, l'application va afficher un message informative pour l'utilisateur pour le informer de saisir les info de colonne cible.

Après avoir rempli les informations nécessaires, il faut cliquer sur le bouton "Submit" pour enregistrer les informations.

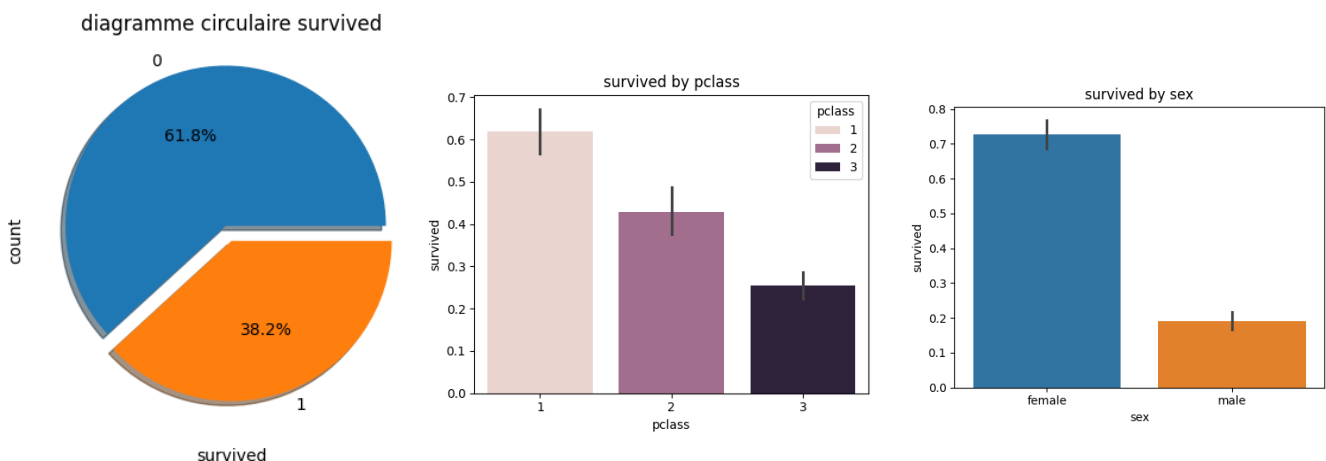
Et l'application va charger le jeu de données et le manipuler utilisant la bibliothèque de Python Pandas.

Visualisation Initiale des Données :

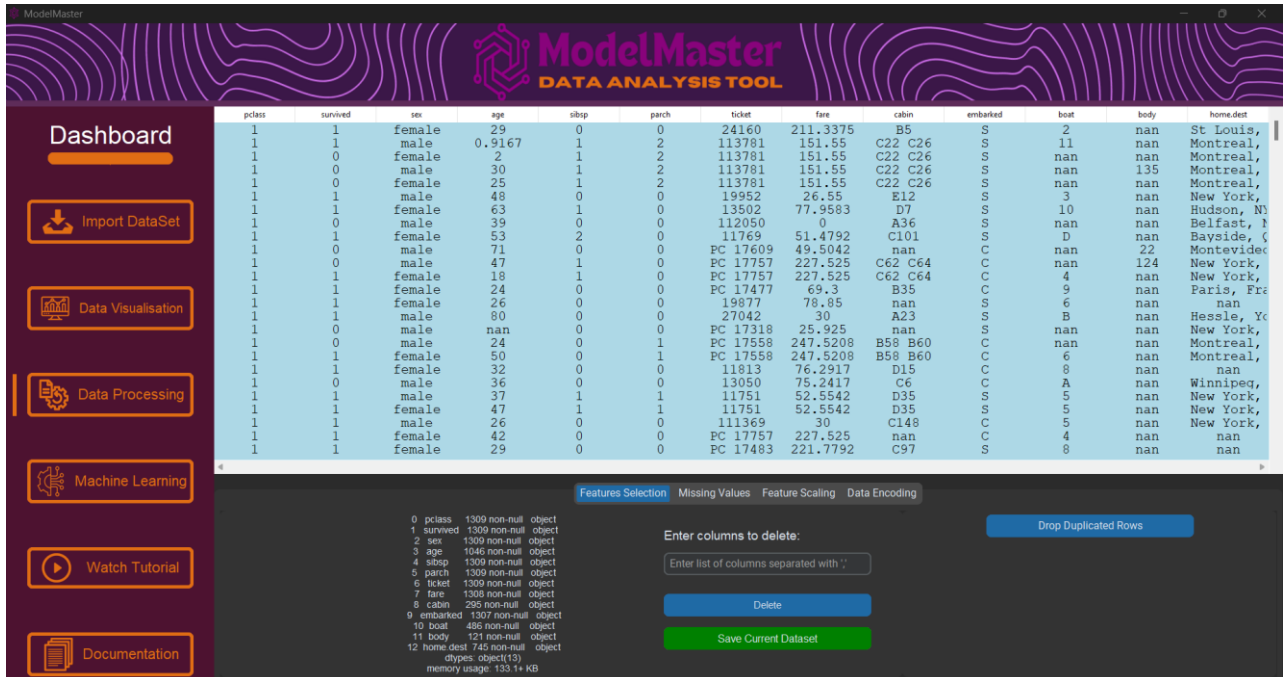


L'interface de Visualisation Initiale des Données a été conçue pour offrir aux utilisateurs une première exploration visuelle de leur ensemble de données. Après avoir sélectionné la colonne cible lors de l'importation, cette section propose des graphiques illustratifs mettant en lumière des informations cruciales. En particulier, des diagrammes spécifiques à la colonne cible choisie sont affichés, fournissant une représentation graphique claire et concise des tendances et des distributions.

Cette interface permet ainsi aux utilisateurs de mieux comprendre la nature de leur variable cible, qu'elle soit catégorielle ou continue. Les diagrammes offrent une vue d'ensemble rapide des patterns et des relations initiales au sein des données, fournissant des insights précieux dès les premières étapes de l'analyse.



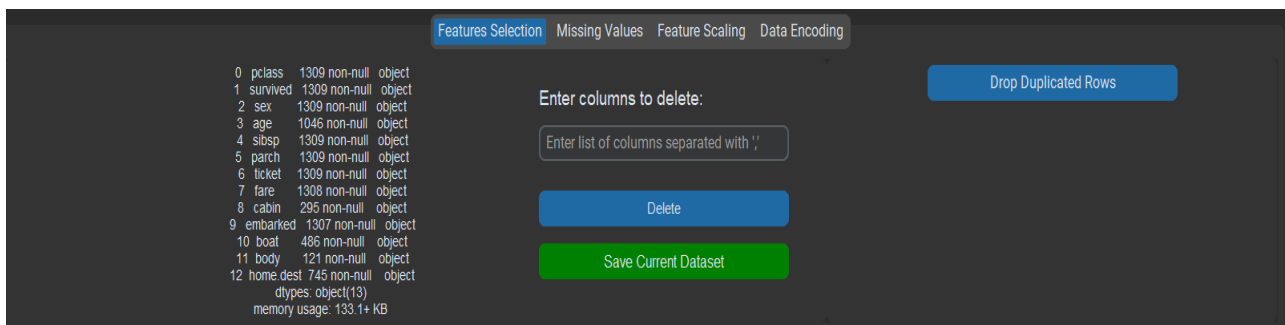
Le traitement des données :



The screenshot shows the ModelMaster DATA ANALYSIS TOOL interface. On the left is a sidebar with navigation buttons: Dashboard, Import DataSet, Data Visualisation, Data Processing, Machine Learning, Watch Tutorial, and Documentation. The main area displays a data table with columns: pclass, survived, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat, body, and home.dest. Below the table, there are tabs for Features Selection, Missing Values, Feature Scaling, and Data Encoding. The Features Selection tab is active, showing a list of features with their data types and a button to 'Drop Duplicated Rows'. There is also a section for 'Enter columns to delete:' with a text input field and buttons for 'Delete' and 'Save Current Dataset'.

L'interface de Traitement des Données offre une plateforme interactive pour la préparation et la transformation des données, mettant en avant une visibilité directe sur le contenu du dataset. Au-dessous de la visualisation du dataset, cinq sections dédiées à différentes fonctions sont présentées :

La première section c'est "Feature selection", qui permet à l'utilisateur de supprimer les colonnes dans le dataset par un clic sur le bouton "delete" après l'utilisateur spécifier les colonnes à supprimer et aussi permet de supprimer les lignes doubles par un clic sur le bouton "Drop Duplicated Row".



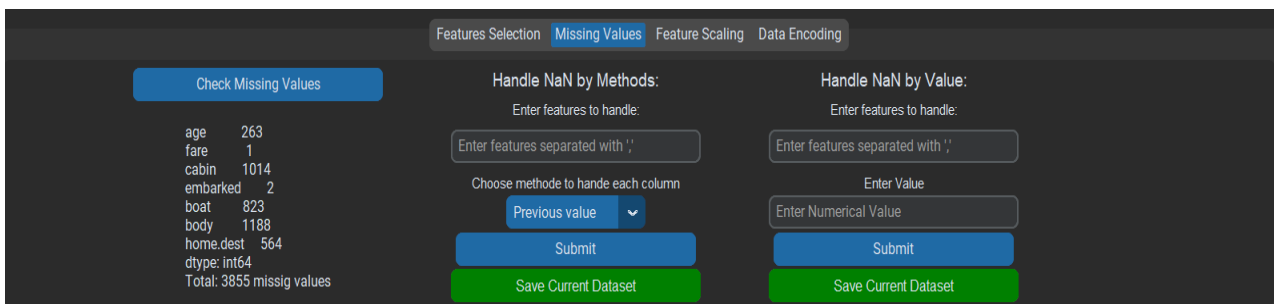
This screenshot shows the 'Features Selection' tab in the ModelMaster DATA ANALYSIS TOOL. It displays a list of features with their data types and a button to 'Drop Duplicated Rows'. There is also a section for 'Enter columns to delete:' with a text input field and buttons for 'Delete' and 'Save Current Dataset'.

Cette section fournit des informations détaillées sur chaque feature, aidant les utilisateurs à prendre des décisions informées qui permettent à l'utilisateur de supprimer les features non pertinentes avec la colonne cible (Target) par l'entrée d'un ou plusieurs noms des colonnes séparées par ',' dans le champ d'écriture. En cliquant sur le bouton "Delete", le processus de suppression est immédiatement initié, offrant une flexibilité significative dans la gestion des caractéristiques du dataset.

La 2eme section c'est « Missing Values », propose une solution efficace pour gérer les valeurs manquantes (NaN) au sein du dataset. En cliquant sur le bouton de « check Missing values », les utilisateurs peuvent vérifier rapidement la présence de valeurs manquantes dans leur ensemble de données. Deux options de gestion des valeurs manquantes sont fournies, offrant une flexibilité adaptée aux besoins spécifiques de l'utilisateur.

La première option permet à l'utilisateur de spécifier le nom de la colonne concernée et de choisir parmi différents types de remplissage pour les valeurs manquantes. Ces options comprennent la moyenne, le minimum, le maximum, la valeur précédente, la suppression de toute la colonne, entre autres. Cette approche offre un contrôle granulaire sur la manière dont les valeurs manquantes sont traitées, permettant une personnalisation précise en fonction de la nature des données.

La deuxième option consiste à saisir le nom de la colonne dans une case dédiée, suivi de la valeur avec laquelle l'utilisateur souhaite remplir les données manquantes. Cette flexibilité permet aux utilisateurs de choisir manuellement la valeur de remplacement en fonction de leurs connaissances spécifiques du dataset.



The screenshot shows the 'Missing Values' section of the ModelMaster interface. It features a 'Check Missing Values' button on the left, which displays a list of features with their respective counts of missing values: age (263), fare (1), cabin (1014), embarked (2), boat (823), body (1188), home.dest (564), dtype: int64, and a total of 3855 missing values. To the right, there are two panels for handling NaN values. The first panel, 'Handle NaN by Methods:', allows users to enter features to handle, choose a method (currently 'Previous value' is selected), and click 'Submit' or 'Save Current Dataset'. The second panel, 'Handle NaN by Value:', allows users to enter features to handle, enter a numerical value, and click 'Submit' or 'Save Current Dataset'.

La 3ième section c'est « Data Encoding », se concentre sur la transformation des données catégorielles, simplifiant ainsi le processus d'encodage. L'interface propose une case où l'utilisateur peut saisir le nom de la colonne qu'il souhaite encoder.

Après avoir spécifié la colonne, deux cases supplémentaires sont mises à disposition. La première permet à l'utilisateur de saisir la valeur qu'il souhaite encoder (par exemple, "male"). La seconde case permet à l'utilisateur d'entrer la valeur souhaitée pour le remplacement (par exemple, "0"). En cliquant sur le bouton "Submit", l'application procède automatiquement au remplacement de toutes les occurrences de "male" par "0" dans la colonne spécifiée.

Features Selection Missing Values Feature Scaling **Data Encoding**

Data Encoding:
Replace categorical values within a specific column with other values
Enter desired feature to perform the encoding:

Enter the categorical value and the value to be replaced with

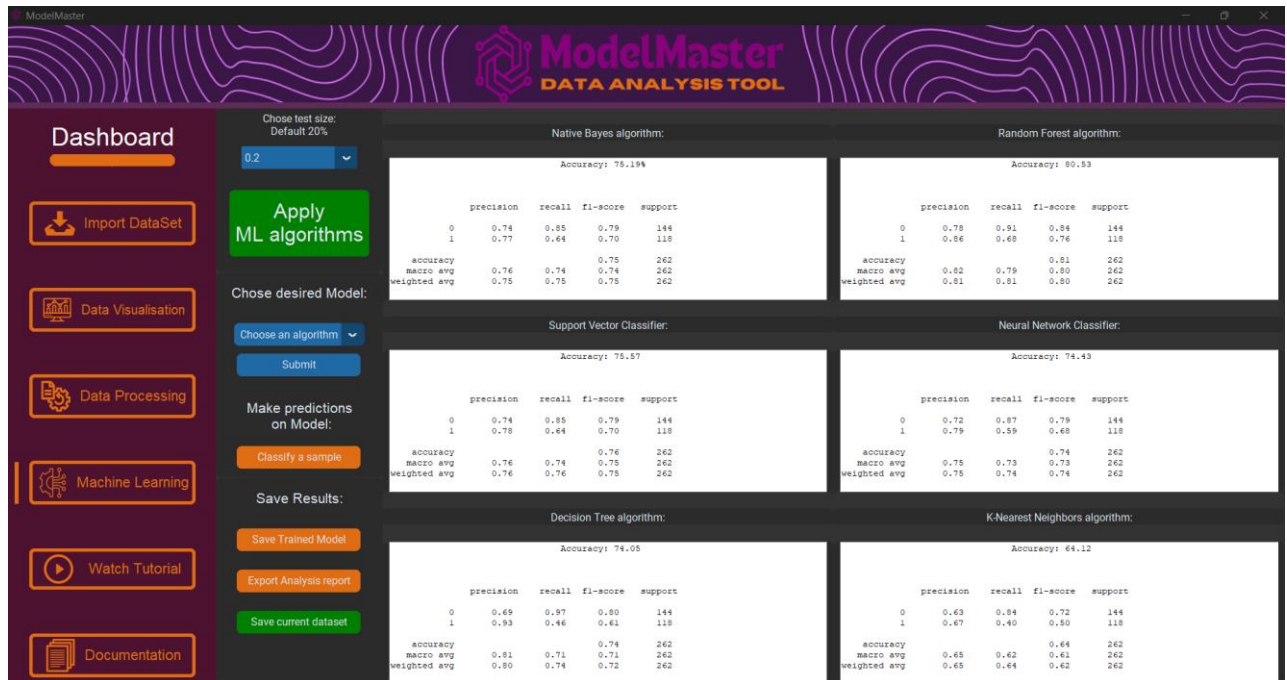
La 4ième et la dernière section c'est « Feature Scaling », met en avant plusieurs options cruciales pour normaliser et standardiser les données, contribuant ainsi à garantir la cohérence et la comparabilité des différentes caractéristiques du dataset.

Features Selection Missing Values **Feature Scaling** Data Encoding

Data Normalization:
Normalization scales the numerical values in the dataset to a range between 0 and 1.
Enter features to Normalize:

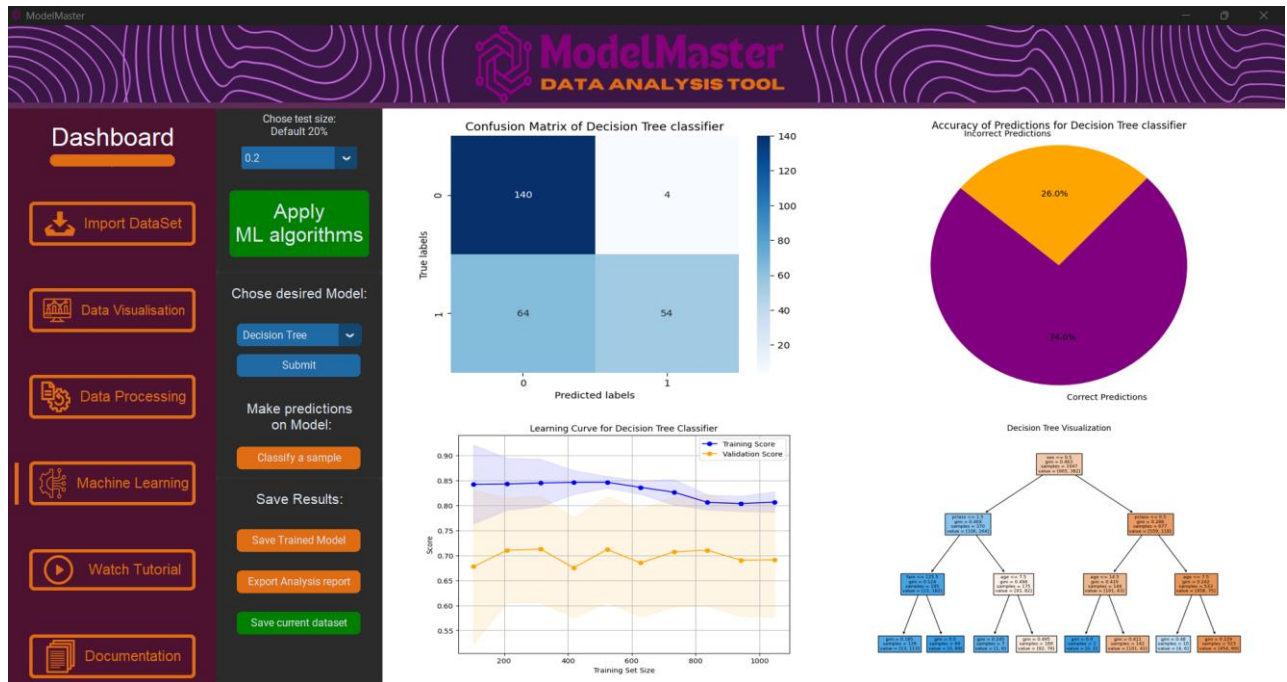
Data Standardization:
Standardization scales the numerical features to have a mean of 0 and a standard deviation of 1.
Enter features to Standardize:

ML Algorithmme et la validation du modèle :



L'interface de "Machine Learning" offre une étape cruciale dans le processus d'analyse, permettant aux utilisateurs d'appliquer différents algorithmes de machine learning sur leurs données. Lorsqu'on clique sur le bouton "Apply ML Algorithm", l'application entreprend une évaluation approfondie en utilisant le plus populaire modèles distincts : Linear Regression, Naive Bayes, Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) et Neural Network. Les résultats obtenus sont présentés sous la forme d'une évaluation de la performance du modèle, détaillant la précision (Accuracy) ainsi que d'autres métriques importantes soit pour les problèmes de classification ou régression.

Ces résultats détaillés offrent une compréhension approfondie de la performance de chaque modèle, mettant en avant des métriques telles que la précision, le rappel, et le score F1. Ils fournissent ainsi des informations essentielles pour aider les utilisateurs à choisir le modèle le plus approprié en fonction de leurs besoins spécifiques et des caractéristiques de leurs données. Cette interface de "Machine Learning" offre donc une évaluation robuste et informative des modèles appliqués.



Chose desired Model:

Decision Tree

Submit

Make predictions on Model:

Classify a sample

Save Results:

Save Trained Model

Export Analysis report

Save current dataset

Après avoir appliqué les modèles de machine learning, l'interface offre une fonctionnalité supplémentaire permettant aux utilisateurs de choisir l'un des modèles à partir d'une liste déroulante. Une fois le modèle sélectionné, en cliquant sur le bouton "Submit", l'application génère des courbes graphiques pour visualiser plus en détail les résultats spécifiques à ce modèle.

Il y a aussi au 4 autre fonctionnalités, ce sont : la 1ere fonctionnalité c'est pour predire la resultat d'un échantillon n'est pas existé dans le dataset, le 2eme pour enregistrer le modèle sélectionné d'un algorithme, le 3eme fonctionnalité pour exporter le résultat du modèle sélectionné sous forme une fichier PDF, le 4eme fonctionnalité pour enregistrer le dataset actuel.

Input fields for prediction:

pclass: 1, sex: 0, age: 1, fare: 0

Predict

Predicted class: 1



Documentation et Tutorial :

Dans la partie de documentation et tutorial, pour la documentation on a posé un fichier PDF contient sur toutes les informations sur l'application, et pour le tutorial on a posé un vidéo explicative détaillé de l'application.

