|  |
| --- |
| Dorset college |
| AI Mini Project |
| Report |

|  |
| --- |
| MULLER Julie  [Date] |

Table des matières

[Instruction 1](#_Toc151108819)

# Instruction

The mini-project has to be done individually.

Solve any two of the four questions below (25Marks each)

Write the python code, conduct hyper-parameter optimization, and provide the best solution.

* Upload the code to GitHub. This should include the following:
  + The python notebook that should contain the following.:
    - Perform exploratory data analysis.
    - Clean the data if needed, i.e., remove or fill missing values etc.
    - Split the data into train-test or train-val-test.
    - Build the model.
    - Train the model and show performance improvement by hyper-parameter modification.
    - Show the evaluation of the model in the test set.
    - Save the model.
    - The final saved model.
* Write a report (2pages/problem) that involves the following:
  + Link to the GitHub page
  + description of the algorithm used for each of the solution.
  + how the model performance was improved
  + show graphs of the final model training and evaluation.
  + The report should be uploaded to Moodle in PDF format.

There are 4 questions provided below. Each problem needs to be solved using provided dataset and proposed method, you can use other method.

Solve any 2 of the following 4 based on the above instructions

* Predict whether a credit card application will be approved or denied based on various features such as income, age, and credit score? Use the Credit Card Approval dataset, which contains information on credit card applications and their associated outcomes, and apply Random Forest to predict whether an application will be approved or denied.
* Classify images of flowers based on their features such as the petal length, petal width, sepal length, and sepal width? Use the Iris Flowers dataset, which contains information on 150 iris flowers belonging to three different species, and apply SVM or random forest to classify the flowers.
* Predict the survival of passengers on the Titanic based on their age, sex, class, and other features? Use the Titanic dataset, which contains information about passengers on the Titanic, and apply logistic regression to predict survival.
* Predict whether a customer will churn (cancel) their subscription to a service based on their usage patterns, demographics, and other features? Use the Telco Customer Churn dataset, which contains information on customers of a telecommunications company and their associated churn status, and apply any suitable method to predict whether a customer is likely to churn or not.

# Telecom customer churn prediction

Predict whether a customer will churn (cancel) their subscription to a service based on their usage patterns, demographics, and other features? Use the Telco Customer Churn dataset, which contains information on customers of a telecommunications company and their associated churn status and apply any suitable method to predict whether a customer is likely to churn or not.

## The libraries

In order to solve this problem we will use the following libraries :

import pandas as pd # data processing

import seaborn as sns # for plots

import matplotlib.pyplot as plt # for plots

from sklearn import metrics #for the accuracy

sns.set(style = 'white')

## The dataset

### Data cleaning

##### Outlier detection

##### Imputation

##### Feature engineering

### Feature selection

## The different models

### Logistic regression

### Random forest

### Vector support machine

1. Add More Data

Not possible cause dataset imposé

1. Treat Missing and Outlier Values

No outlier values or missing ones detected

1. Feature Engineering
   1. Feature transformation
   2. One-hot encoding
2. Feature Selection
3. Multiple Algorithms
4. Algorithm Tuning
5. Ensemble Methods
6. Cross Validation

* Write a report (2pages/problem) that involves the following:
  + Link to the Github page
  + description of the algorithm used for each of the solution
  + how the model performance was improved
  + show graphs of the final model training and evaluation.
  + The report should be uploaded to moodle in**PDF format**