Task 1: Data set

* Choose the data set

Task 2:

* Define the problem statement or state the situation because you want to make sure that the machine learning solution you are providing is aligned with your client’s needs.
* Identify weather the project tackles a problem of regression, clustering or classification?

Task 3:

* Brief description of the data set and a summary of its attributes

Task 4:

Perform data preprocessing tasks, including handling missing values, scaling features, encoding categorical variables.

Actions taken for data cleaning and feature engineering.

* Drop unrequired column
* Drop goalkeeper
* Choose labels and features
* Normalize or standardize …

Task 5:

Apply techniques of data exploration to understand the data set, identify relevant features.

Key Findings and Insights, which synthesizes the results of Exploratory Data Analysis in an insightful and actionable manner. Visualization Charts

Task 6:

Apply multiple machine learning algorithms

If you choose to use two ML algorithms (e.g. Decision tree classifier, random forest classifier,

logistic regression, and so on) not taught in class you get bonus points.

* Implement KNN
* Learn how Random Forest works and implement it
* Learn how SVM works and implement it
* Learn how Logistic Regression (Softmax) works and implement it

Task 7:

Implement and compare different metrics to evaluate the quality of the model. Discuss the results. You may use various plots also.

Task 8:

Clearly present the findings, insights and recommendations

Compare the results obtained from the different applied methods and discuss the limitation of the models before presenting areas for future improvement