EPITA - International Programs

AI Project Methodology

2022-2023

**Project**

**Goals**

The goal of this project is to apply some concepts & tools seen in this course :

1/Framing methodology

2/Technical methodologyorganized into 3 parts :

●Part 1 : Building Classical ML project with respect to basic ML Coding best practices

●Part 2 : Integrate MLFlow to your project

●Part 3 : Integrate ML Interpretability to your project

**1/Functional Part : Define/Frame your Project (think our strategy)**

**Problem**: Predicting the outcome of a basketball game based on team statistics.

**Goal**: To build a machine learning model that accurately predicts the outcome of basketball games based on team statistics.

Next, let's identify the customers/users who will benefit from our project:

**Customers/Users**: Basketball fans, sports betting companies, and coaches/managers who want to analyze their team's performance and make data-driven decisions.

Now, let's determine the datasets we will use for our project and their source:

**Datasets**: NBA game data from the 2019-2020 regular season. The data includes game statistics for each team, such as points scored, rebounds, assists, turnovers, and shooting percentages. We obtained ours from Kaggle.

After identifying the problem, goals, customers/users, and datasets, it's important to define the project management framework we will use. There are several project management frameworks to choose from, such as Agile, Waterfall, and Scrum. It's up to our team to decide which framework works best for our project.

**Project Management Framework**: Scrum. Scrum is an Agile framework that emphasizes collaboration, flexibility, and continuous improvement. It's well-suited for projects with rapidly changing requirements and a focus on delivering value to customers.

Now, let's define the roles and organization of our team:

**Team Members**: Rayane and Remi.

**Roles**:

* Rayane will focus on data preparation and cleaning, as well as model training and evaluation.
* Remi will be responsible for feature engineering, model selection, and tuning.

**Organization**:

* Both team members will work closely together to define the problem, choose the appropriate model, and interpret the results.
* Rayane and Remi will meet weekly to discuss progress, address any issues, and plan for the upcoming week.
* Both team members will have access to the same code repository and work on separate branches, merging their code regularly.

With these elements defined, we can now begin working on Part 1 of our project: Building a Classical ML project with respect to basic ML coding best practices. This involves the following steps:

1. Data Preparation:

* Obtain and clean the data from the chosen source.
* Perform exploratory data analysis to identify any missing or inconsistent data.
* Transform the data into a format suitable for training the machine learning model.

1. Feature Engineering:

* Select relevant features that are likely to have a strong correlation with the outcome of basketball games.
* Create new features that combine multiple variables to provide more predictive power.

1. Model Selection:

* Choose an appropriate machine learning algorithm to predict the outcome of basketball games.
* Split the data into training and testing sets to evaluate the model's performance.

1. Model Training:

* Train the selected machine learning algorithm on the training data.
* Evaluate the model's performance on the testing data.

1. Model Evaluation and Interpretation:

* Interpret the results of the trained model to identify which features are most important in predicting the outcome of basketball games.
* Refine the model by adjusting hyperparameters and repeating the training and evaluation steps.

By following these steps, we can build a machine learning model that accurately predicts the outcome of basketball games based on team statistics, using basic ML coding best practices

Here is the accuracy we obtained:

Text

Description automatically generated with low confidence

As a student, it's important to schedule our ML pipeline so that we can make sure our code runs in the correct order and everything is executed as we intend it to be. One solution is to use a tool like Luigi or Airflow, which allows us to specify the dependencies between tasks and run them in parallel. Another solution is to use a simple script or makefile that runs each script in order, making sure that the output of one script is the input of the next script. By scheduling our ML pipeline, we can ensure that we are making the most efficient use of our time and computing resources, and that our code is always up-to-date with the latest data.