Achieving a ML Proof-of-Concept

2024-2025

Project Guidelines

Gianluca Quercini

ALBERTSCHOOL

Guidelines

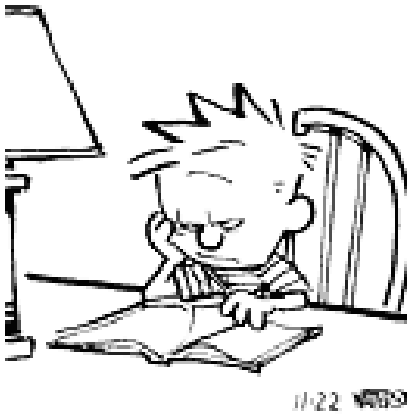
- **Project Objective:** The aim is to develop a proof-of-concept to assess the feasibility of using machine learning techniques to make predictions on a dataset. The project consists of the following steps: **data collection, data cleaning**, **feature engineering**, machine learning model **training** and **testing** and **user interface** development.

- **Data collection and cleaning:** You must gather data from **APIs** and/or **scraping**. The data will likely come from several sources, your task will be to **clean** them and remove anomalies (duplicates, erroneous and inconsistent data) as much as possible.

- **Machine learning models:** You must train and test **at least three machine learning models**. You must compare the three models using the appropriate accuracy measures. You'll need to use the appropriate **feature engineering** techniques.

- **Groups:** This project is to be completed in groups. You should choose by now a **team leader**. The team leader will have the responsibility of cloning the initial GitHub repository.

- **Evaluation:** Four milestones will be evaluated throughout the course (more on this later).



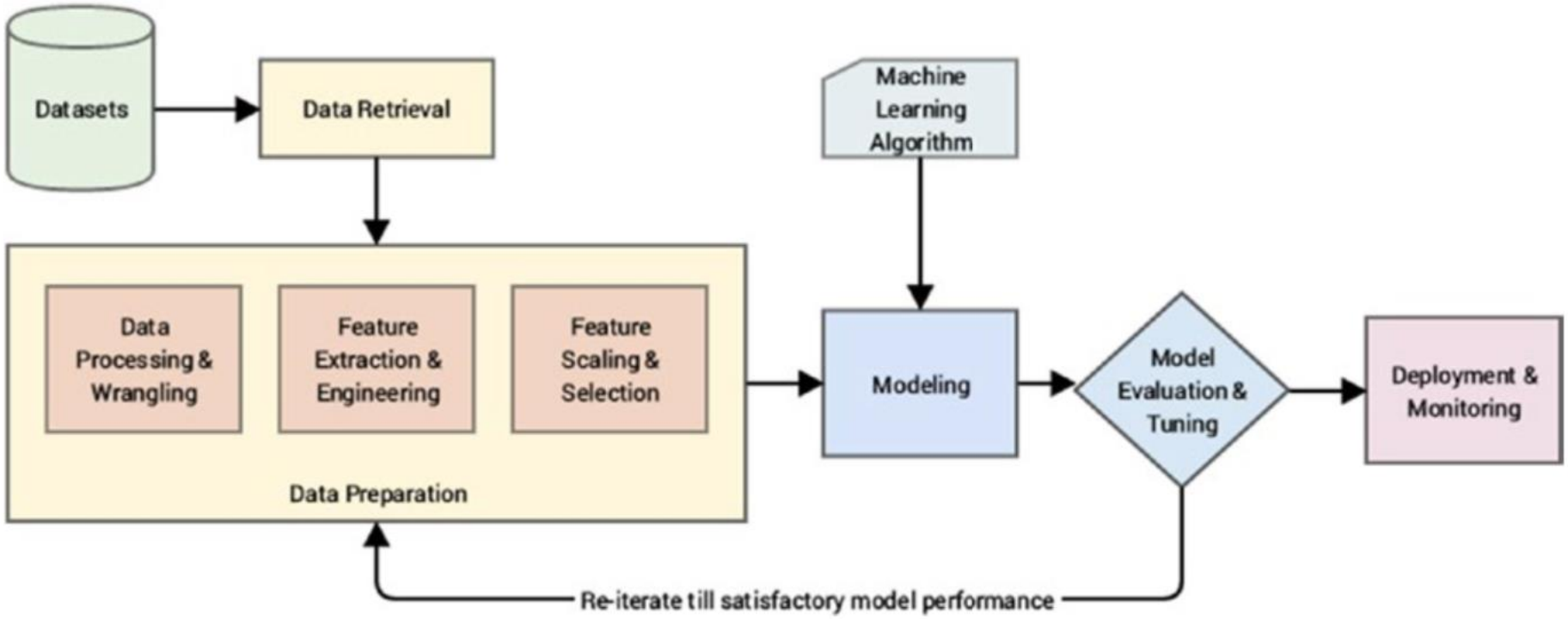
Follow the best practices for data projects to ensure success !

ML stages ➡ key steps

1. Look at the big picture.
2. Get the data.
3. Discover and visualize the data.
4. Prepare the data for Machine Learning algorithms.
5. Select a model and train it.
6. Fine-tune your model(s)
7. Present your solution.
8. Launch, monitor, and maintain your system.

ML Stages DALL-E generated

Let's go back to ML stages → ML flow



https://web2.qatar.cmu.edu/~gdicaro/15288/

Evaluation

- **Project setup (10/04, 4:30 PM):** Setup of a GitHub project and raw dataset.
- **Intermediate evaluation (30/04, 1:15 PM): Cleaned dataset, test of at least one machine learning model.**
- **Git knowledge check (06/05 1:30 PM):** Quick knowledge test on Git (QCM on Evalmee, 15 minutes).
- **Final evaluation (16/05):** Oral presentation (15 minutes, including questions). **All files, including the PowerPoint, must be submitted before 15/05 11:59 PM**.

Coefficients

- **Project setup:** 5%
- **Intermediate evaluation:** 20%
- **Git knowledge check:** 5%
- **Final evaluation:** 70%

Evaluation - Rules

- **Each inappropriate behavior** during the sessions (leaving the room for a long time, working on other assignments, using mobile phones…) **will be penalized with -1 points on the overall grade**.

- **Failure to comply with a deadline** will result in a **grade of 0** on the corresponding evaluation item.

- **You must attend at least 75%** of the sessions to receive a grade at the intermediate and final evaluation (otherwise, the grade will be 0).

- The **grade of the group is applied to each member, unless a member is seen working much less (or with a high number of absences).**