

Machine Learning Project Checklist



You have been hired as a new Data Scientist and you have an exciting project to work on! How should you go about it?

In this lesson, we are going to go through a checklist and talk about some best practices that you should consider adopting when working on an end-to-end ML project.

A checklist to guide you through your machine learning projects

1. Frame the problem and look at the big picture:

- Understand the problem, both formally and informally
- Figure out the right questions to ask and how to frame them
- Understand the assumptions based on domain knowledge

2. Get the data: Do NOT forget about data privacy and compliance here, they are of paramount importance! Ask questions and engage with stakeholders, if needed.

3. **Explore** the data and **extract insights**:

- **Summarize** the data: find the type of variables or map out the underlying data structure, find correlations among variables, identify the most important variables, check for missing values and mistakes in the data etc.
- **Visualize** the data to take a broad look at patterns, trends, anomalies, and outliers. Use data summarization and data visualization techniques to understand the story the data is telling you.

4. **Start simple**: Begin with a very simplistic model, like linear or logistic regression, with minimal and prominent features (directly observed and reported features). This will allow you to gain a good familiarity with the problem at hand and also set the right direction for the next steps.

5. **More Feature Engineering**: Prepare the data to extract the more intricate data patterns. Combine and modify existing features to create new features.

6. Explore **many different models** and short-list the best ones based on comparative evaluation, e.g., compare RMSE or ROC-AUC scores for different models.

7. **Fine-tune** the parameters of your models and consider combining them for the best results.

8. **Present** your solution to the stakeholders in a **simple, engaging, and visually appealing manner**. Use the **art of story-telling**, it works wonders!

- Remember to **tailor your presentation** based on the technical level of your target audience. For example, when presenting to non-technical stakeholders, remember to convey key insights without using heavy technical jargon. They are likely not going to be interested in hearing about all the cool ML techniques you adopted, but rather on end results and key insights.

8. If the scope of your project is more than just extracting and presenting insights from data, proceed with launching, monitoring, and maintaining your system.

Of course, this checklist is just a reference for getting started. Once you start working on real projects, adapt, improvise, and at the end of each project reflect on the takeaways, ***learning from mistakes is essential!***

Note: We are going to go into the technical details of all these steps, and their sub-steps, in the “*Project Lessons*”; the purpose of this checklist is to serve as a very high-level guideline or best practices reminder.

With this, we are finally ready to move on to our project and get our hands dirty! 📖👊