ROCKET SCIENCE

PHASE 3

The feasibility of a Machine learning project

Now that we know how to detect projects with high potential, we can move on and learn how to evaluate the feasibility of a project. Note that having a project with high potential doesn't mean that it is a feasible project.

Feasibility means that a project is technically doable. Normally it requires deep technical knowledge with good industry experience to know if a project is really feasible. And due to the uncertainty in Machine Learning, a feasible project on paper can end up being a not possible project during development.

This framework aims to simplify your decision regarding how feasible a project is. We recommend validating the feasibility by a specialist in Machine Learning that will guide you around the feasibility of your projects.

Question to ask when analyzing the feasibility

Is the data available internally or does it need to be acquired externally?

Very frequently, what limits a project is the availability of data. Companies and leaders end up discovering that internal data is not enough to conduct a specific project, and that no third party can provide this data, or that buying third-party data is too expensive. The most popular use case is in marketing, where companies want to conduct Machine Learning projects on their customer data (like predicting customer's behavior) and they realise that they are missing data about their customers (such as age, sex, and any socio-demographics). When turning to third party providers, this can become very costly, especially when you have millions of customers with many characteristics desired.

This reflects if the company collects the data related to a specific project. Sometimes collecting data is not possible for marketing or ethical reasons, and sometimes it can be time-consuming as it can take a long time before having a decent amount collected.

> So, the availability of the data can play a critical role in the feasibility analysis.

Does the project have existing use cases on the web or from other companies?

If a project is already documented on the web as a use case, this means that it is probably a feasible project. For example, when you google "customer churn prediction", you will find many use cases on the web and even companies explaining how they implemented this Machine Learning project in their organisation. As well by attending conferences in your industry or AI, you might hear about use cases related to your potential projects that explain how it has been implemented.

All these are signs that your project might be feasible. But there is no guarantee that your project will be feasible technically in reality as every company and project are unique. However, it is still a good indicator of the feasibility.

How hard would it be to deploy and maintain the model in production?

Another factor that can play a certain role in the feasibility analysis is how this model will be deployed. And how costly it is to maintain in production. For example, a telecommunication company would like to create a Machine Learning model that can detect any anomaly on their network data. This project requires creating ingestion pipelines in production. Then the pipelines are streamed with the machine learning model deployed. Creating such streams and pipelines for all network data is very costly in terms of hardware infrastructure, but streams also require efforts to be maintained and updated.

Defining how hard it is to deploy and maintain can get quite technical which might require you to speak with a specialist (a Machine Learning Engineer or MLOps specialist).

Scenarios

A retail company wants to analyse if a customer churn project is feasible. This company has assessed that a customer churn prediction project is highly valuable.

First, the company looked at the data collected and available. They realized that they have transactional data about the customers, as well they have basic information about their customers, like the address. Another data could be available, and that is the data about calls to customer support, including duration of calls, number of calls per month, etc. Having access to additional data about customer socio-demographics would have been great to support the project, but they don't want to ask their customers to provide this information and are not willing to purchase it from third-party providers. All this makes the project quite feasible technically from a data perspective.

One of the leaders decided to google "customer churn prediction", and realized that churn prediction is a well-documented project with many use cases available online. This makes the project feasible from a tracking record perspective.

From deployment and production perspective, the models could be deployed as a simple script against the databases, and this would perform pretty well. Other more sophisticated deployments are possible like deploying in Microservice which would allow integration with other tools and software. This makes the project feasible from a deployment perspective with average to low complexity.

Overall performance

Data	Enough to start first POC/Prototype
Tracking record	Highly available
Deployment and maintenance in production	Average to low complexity

What do I do if the project is not really feasible?

Well, a project that is qualified as not feasible can still be conducted depending on what aspect is not feasible. If the project is not feasible due to a lack of data, the project can be delayed waiting to collect enough data. If the project is not feasible due to a lack of tracking record (no existing use case found), the project can be conducted in an R&D mode. If a project is not feasible from a production perspective, the company can start implementing on a smaller scale or in a simulation environment. Working with Machine Learning Architects can help design an optimized and engineered environment for production that will lower production costs and efforts.