

RECIPE FOR A SUCCESSFUL DATA SCIENCE PROJECT

fastdatascience.com

PRE-PROJECT

DEFINE THE GOALS

Discuss with stakeholders.
Core business question?
What are the goals of the project?
Define initial hypothesis.

INITIAL EXPLORATION

Play with some basic machine learning models such as Scikit-Learn.
Confirm if project is possible.
Interface with stakeholders.

DATA COLLECTION

Collect a small sample of data for exploration.

AGREE SCOPE AND COMMERCIALS

Involve stakeholders fully.
Define KPIs and acceptance criteria.
It must be clear what the ROI is if the project proceeds.
Stakeholder buy-in is essential.

DURING THE PROJECT

INVESTIGATION STAGE

Set up testing framework.
Develop and test hundreds of models.
Experiment with hyperparameters.
Show to stakeholders and ask for feedback.

QUALITY CONTROL

Rigorous stress testing of model.
Evaluate for bias, changing conditions, non-stationarity.
Get stakeholder signoff.

MODEL REFINEMENT

Choose a model architecture.
Choose the best configuration.
Iron out the bugs.

DEPLOYMENT

Involve stakeholder DevOps.
Set up testing framework and monitoring.
Batch job to retrain regularly?

POST-PROJECT

HANDOVER

Follow up with stakeholder and check that they are satisfied with project.
Knowledge transfer to organisation: presentation, Word documents, executive summary.

SHARE

Don't forget to share the success internally.
If permitted, share publicly.

SAY THANKS

Thank the stakeholder for their cooperation.
Hope to work together again.