Project Success Prediction This is a demonstration of predicting which projects are eventually judged successful by the World Bank, using their data.

Why The reason why I have done this is to start to understand how to apply machine learning to portfolios of projects. Predicting which projects are more likely to succeed would have a high return both in terms of reduced costs and benefits accruing from more successful projects. This has been in a low-code environment to show how easily much of this can be understood and applied by normal portfolio and project managers.

What we would be trying to achieve on behalf of a client like the World Bank, or any client with a lot of projects or a lot of work-packages, is to train a machine-learning model on their archive of projects, and which will then predict which projects are most likely to succeed and which to fail.

Data We started with the World Bank because they have opened their data on 12000 projects from the last fifty years. Each project has been given a final class at the end of the Project: Highly Satisfactory, Satisfactory, Unsatisfactory etc. This is what we care about. We can train a model using supervised machine learning to predict this classification for projects that the model has not seen before.

Outcome For projects that the model had not seen, the model achieved a classification accuracy of 89% . i.e. for every 100 predictions it made, it got 89 right.

Conclusion for our business: This experiment shows a good level of success. We believe it is worth exploring further the applications of machine learning for advising companies with large portfolios. There are already many tools that can do this, and this is just one project use case of several which are already possible with the new methods.

How we did this:

- **Extract, transform and load the data:** We brought the 12000 projects in, and tidied up the data. As well as the final status of the project, each project has about 20 features such as date, duration, cost, and the results of different interim project reviews. This is the raw material that the model can use to understand what combinations of features are most predictive of eventual success.
- **Train the model:** We applied the project data to several machine-learning algorithms that learn to classify the projects into Highly Satisfactory, Satisfactory, etc. We compared the models and chose the best performing model
- **Apply the model:** Once trained, the above steps do not need to be re-run. The model can instead be periodically applied by a someone who has had a day or two of training. It will predict for an unseen project whether the project is likely to ultimately be satisfactory.