

Machine learning Use Case : Project Success Prediction for large portfolios

What task/decision are you examining? We are examining the way that a portfolio of projects is overseen by a large Institution. We have started with the World Bank as they have data on 12000 projects and which were ultimately successful. The decision we are seeking to improve is in identifying which projects are unlikely to succeed, and whether they should be cancelled, rescoped, supported, or monitored as a result.

Prediction: Predict during project implementation which World Bank projects will be evaluated as Satisfactory after the project completes.

Judgement: The payoffs of being right for the Institution are being able to focus on shutting or intervening in projects that are likely to fail. The impact of false positives is cancelling projects that might otherwise be successful. The impact of false negatives is less serious as it would just lessen the improvement in project coverage

Action: The actions that can be chosen as a result of the judgement would be to:

1. Review a project forecast to fail
2. Strengthen resource on these projects

Outcomes: We would judge whether we are achieving our outcomes by monitoring whether

1. Projects forecast to succeed actually do succeed. For the WB and this dataset, since accuracy achieved is 90%, this should be the target during production.
2. What % projects forecast to fail do fail – anything over 50% would be good

Training: To do this we used a sample of the 12000 previous projects. Each project has been rated as Satisfactory, or Unsatisfactory, Highly Satisfactory etc. after the project concluded by the client. This rating label has been used as the target for Learning. Each project has about 20 features recorded for it, and our model is trained to understand which combinations of these features are most predictive of success. This is a supervised learning approach, where we have ended up training and selecting a Random Forest Model as providing the best results.

Input: Now that we have trained the model, we would need data on currently running projects as per the feature list used in training. These include forecast to completion, mid-project quality reviews and industry area. This could be fed into the model once a month by the portfolio manager.

Feedback: During production, we will need to use measured outcomes along with input data to generate improvements to our predictive algorithm. This can be done semi-annually, re-running and refitting the whole model. A new model would then need to be issued to the portfolio manager.

How will this AI impact on the overall workflow? Regularly predicting project success will focus management attention on the right projects. It will improve forecasting accuracy, and thereby should increase overall project completion. It may also have a lesser effect on early project scoping and project selection. The impact on the portfolio manager would be a day a month. Since there are likely to be project reviews and interventions for failing projects already, it is likely that the time taken on these from model predictions would be substitution of existing tasks rather than new tasks. As this is implemented, there would be a backlog effect and a higher turnover and reallocation of project teams.

