

Dataro's Machine Learning Pipeline



Dataro uses powerful machine learning technology to turn your data into predictions about how donors are likely to behave.

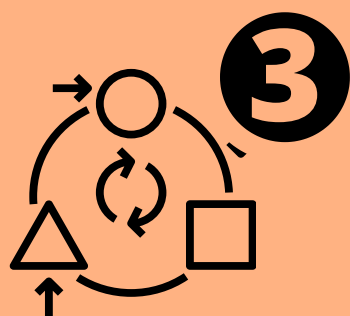
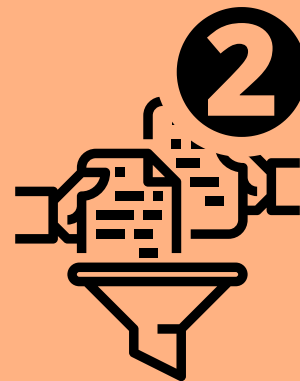


1 Extract Data

We extract data from your CRM using our secure integrations which make use of your system's APIs or inbuilt reporting functions. Our modelling schema is based on donors (Constituents or Contacts), donations (Gifts or Opportunities), campaigns (Assigned Appeals or Campaign Members) and ad-hoc interactions (Actions and Activities). We do not require any personally identifiable information and the data is stored encrypted in the cloud.

2 Prepare Data

Your data is cleaned and normalised to fit our modelling schema. We hunt for and fix common issues (e.g. out-of-range dates, impossibly high or low values) then convert your CRM's data model into our modelling schema. This includes interpreting your data (e.g. identifying appeals, regular gifts and campaign channels.)

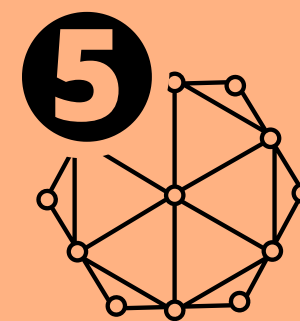


3 Transform Data

We then transform your data into a potential set of over 400 numeric 'features' suitable for machine learning. This could include a donor's age, or the number of days since their last gift. Of these, anywhere from 10-180 features may end up being important in the final model.

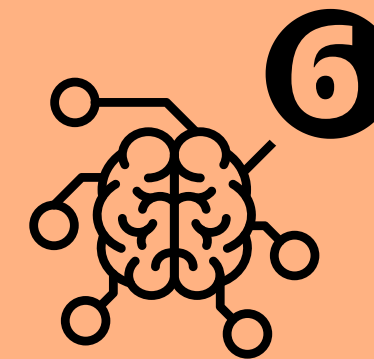
4 Create Timeslices

To train the models, we prepare datasets using a 'time slice' approach. Imagine that we take 'snapshots' of the data at many points in the past and then combine these snapshots together. Our models learn the patterns in this data to predict what happens next and are then tested against withheld datasets.



5 Train Models

We train numerous algorithms and models for every propensity. The algorithms include rule-based methods (e.g. RFM) and simple linear models (e.g. logistic regression) through to gradient boosted models (e.g. XGBoost) and neural networks (Multi Layer Perceptrons a.k.a. 'Deep Learning'). We also test a range of 'hyper-parameters' for each algorithm to find the most appropriate model for your data and the propensity we are interested in (e.g. appeal giving). The more complex models are generally suited to detecting subtle patterns in rich datasets, while the simpler models are more reliable when modelling events that occur more sparsely in the data.



6 Select Model

Each model is evaluated against a range of stringent testing criteria and the most effective model is selected to generate your predictions.

7 Generate Predictions

Here's the fun part. Using our chosen model, we generate 'Propensity Scores' and 'Propensity Ranks' for every donor for every propensity you have selected. 'Propensity Scores' tell you the probability that a donor will take a certain action, like give to an appeal. So a score of 0.59 means they are 59% likely to give. 'Propensity Ranks' tell you how that person compares to everyone else in the dataset. So a rank of 10 means they are the tenth most likely.



8 Load into CRM

We create custom fields in your CRM to store this new data. These fields are queryable so you can create reports and campaign lists easily. We use APIs or upload scripts (depending on your CRM) to load new predictions on a weekly basis -- all this is done automatically without needing to worry about file transfers or an upload process on your end!.