Santander Banking Project

How we turned millions of numbers in millions in profit

Problem to solve

Predict whether a **customer** was going to make a **transaction or not** based on **200 anonymised features**

Project outline

Data cleaning

We checked for inconsistencies or missing values in the dataset (200’000 samples), while transforming our data to have comparable scales over the entire dataset

EDA

All features had standard distributions. There was no single group that carried the majority of the predictive value (everything was important)

Modeling

We tested different models, from “dummy” ones as baselines, to class-standards, to more complex ones

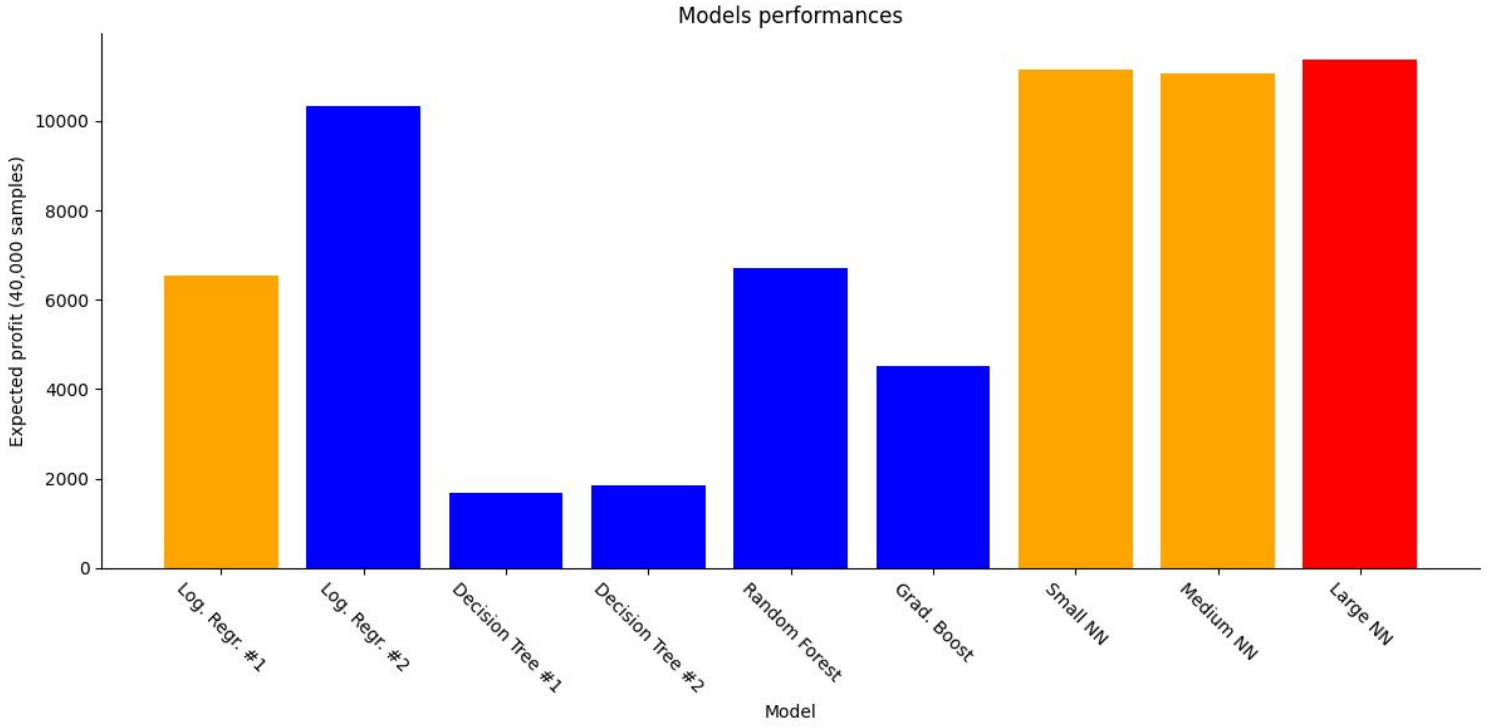
Pipelining

Once decided on the model to use, we made a fully automated, production ready pipeline (can be found in the notebook)

Tested models

Best performance → **Large neural network**

*(**1.50€ increase** in average profit-per-customer over our baseline)*



Real-world metrics

1.50€ profit-per-customer

at **60M** annual customers

⇒ **over 90.000.000 €** in added annual profit

Real-world metrics

Full **production-ready pipeline** can be found in the notebook