

## Machine Learning Lifecycle

Machine Learning Lifecycle consists of 4 steps:

1. Data Gathering
2. Data Preparation / EDA
3. Modelling (Build, Train, Test, Tune)
4. Deployment

While working on our project we covered them all. The first thing we had to do was to gather the needed data and analyse if our project idea was feasible. After collecting all the data we encountered several issues and adjusted our project idea.

The next step was to explore, visualise, merge, and prepare all the data for modelling. We created an EDA report where we documented all our findings and data preparation steps. Also, we merged all the datasets based on the neighbourhood of the Breda they were related to.

Later we started with the modelling part. We tried several models such as ARIMA, VARMA, Gradient Boosting Regressors, and Prophet. We decided to use the last one, as it gave us the best results and was the easiest to implement, evaluate, and tune. While working on this model, we have performed feature engineering by transforming time series into a supervised learning problem. This step helped us to pick the best predictors for our model, which is why we created a separate model for each neighbourhood of Breda. After training the model we evaluated it using metrics such as MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and others. We picked these metrics as we were trying to predict the number of crimes in a neighbourhood, so it was hard to measure the accuracy of the model and we decided to track the errors in the model's predictions. Subsequently, we tuned it by going through different values of the 'change\_prior\_scale' and 'seasonality\_prior\_scale' variables and picked the values with the lowest MSE and RMSE. By tuning the model we managed to decrease the errors in predictions by 20%.

The steps mentioned above made it possible for us to create a high-quality model that predicted the monthly number of crimes in the neighbourhoods of Breda. Last but not least, we deployed our model by creating a web application using a Python library called Streamlit. Here comes the end of the machine learning lifecycle for our project.