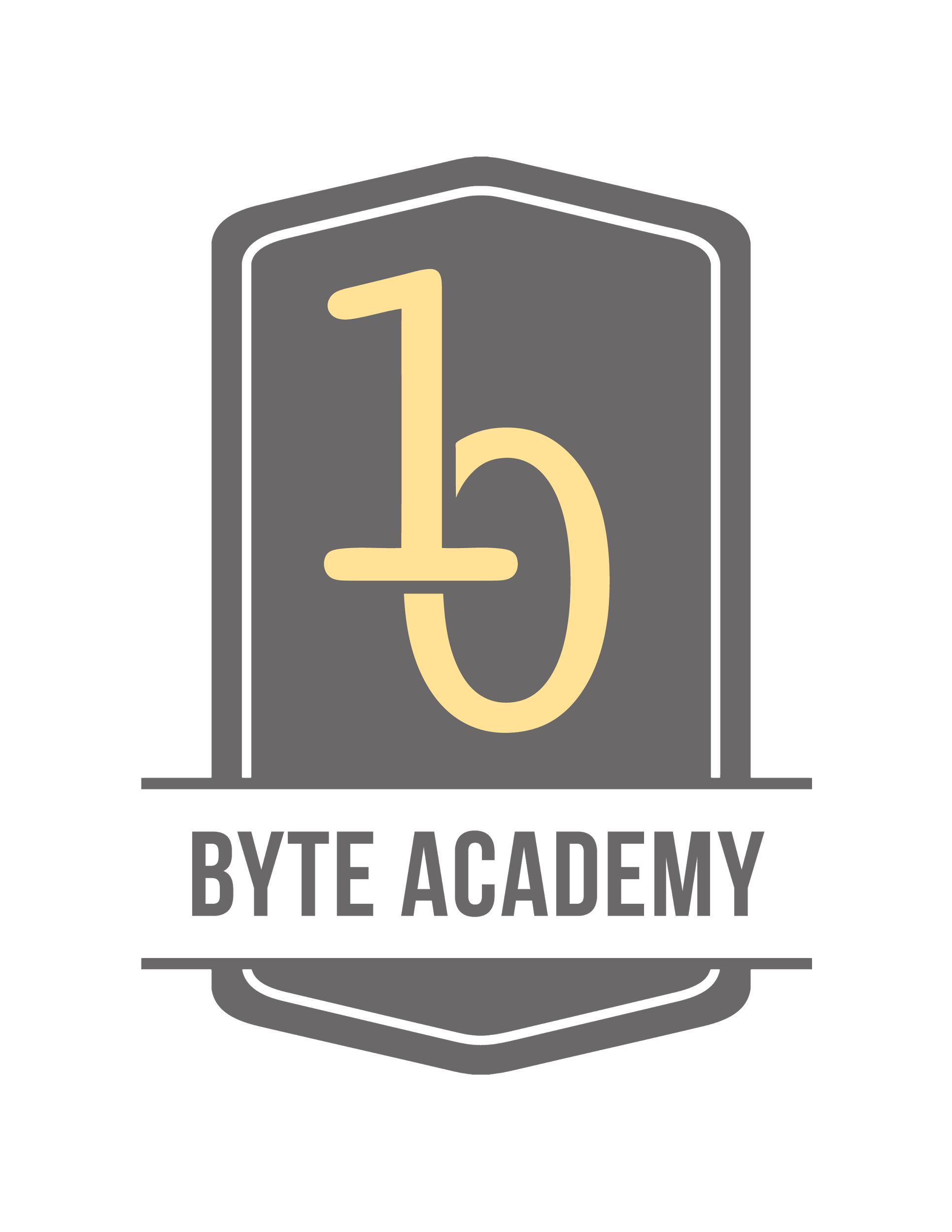
*Project Options for Data Science*

**Predict popularity of listing on NoBroker.com**

Tags: Numerical Prediction, NLP, Real-Estate, Scraping-Required

*Description:*

1) Task: Predict interest level of a rental listing on nobroker.com

2) Target variable: Number of times listing was shortlisted.

3) Attributes: Must be scraped from the site but should include the description

4) Evaluation: Hold out 20-30% of data for testing and minimise the RMSE

**Prediction using Lending Club data**

Tags: Finance, Classification, No-scraping Required

*Description:*

The data on Lending club can be used to predict different targets:

- Predict borrower’s default probability

- Predict probability of a loan being accepted/decline

1) Attributes: Choose most relevant attributes out of 115 potential options

2) Evaluation: Attain highest precision/recall using 20-30% for testing high and/or high AUCROC if predicting accepted/decline. If predicting probability, minimise RMSE.

**Predict mortality of patients in ICU**

Tags: Medical, Health Care, Binary-Classification, No-scraping required

1) Based on a 2012 challenge on Physionet

2) Target Variable: Survival (0), death(1)

3) Evaluation: Attain highest precision/recall using 20-30% for testing high and/or high AUCROC

**Build a recommendation engine (AKA The Netflix Prize)**

Tags: Recommendation, Netflix, No-scraping required

1) Utilise the industry-standard collaborative filtering algorithm.

2) Evaluation: Get the lowest possible Root Mean Squared Error (RMSE)