Stage 3. ML problem formulation – Model and Loss

Supporting materials:

- **Machine Learning: The Basics. Chapter 2.2, 2.3 & 6.2.**

Introduction*

Problem Formulation*

*You can copy-paste/ edit these sections from previous tasks based on received feedback. Alternatively, you can choose completely different problem.

Methods**

- State the number of datapoints, briefly describe the dataset and/or any data preprocessing or cleaning needed.
- If using categorical or ordinal variables, explain how you encode them.
- Describe and explain (why?) your choice of ML model(s)/hypothesis space(s)***, e.g., linear predictors, etc.
- Describe and explain (why?) your choice of loss function(s), e.g., logistic loss.
- Explain the process of model validation - how did you split the data into the training, validation and test sets. What are the sizes of each set and why did you make such design choice.
- If applicable, describe and explain your use of metrics (e.g. accuracy) in addition to loss function (e.g. logistic loss).

** No implementation needed, just describe planned experiments.
*** To get max points, use **at least two different** hypothesis spaces.