CLASS PROJECT

- Team effort, 3-5 people. Teams of one are possible, but prior discussion is needed

- Recommended data source – Kaggle **kaggle**
- Consider necessary compute time!
- Timeline:
  - Oct 10 – teams are formed, dataset selected and registration spreadsheet filled
  - Nov 12,14 – Interim project review. Short 3-min presentation about your progress
  - Nov 30 – Final report is due. Approximately 5 pages + code in a publicly accessible git repository. Link should be included in the report document

CLASS PROJECT REQUIREMENTS

- Explore dataset – correlation between features, their impact on target, independence assumptions
- Try multiple visualization strategies: correlation plots, dimensionality reduction approaches – PCA, UMAP, t-SNE. Explain gained insights
- Explore data using unsupervised learning techniques
- Identify appropriate cross-validation strategy
- Train a simple model first. Use validation set for hyperparameter tuning and/or early stopping. Analyze its performance using cross-validation. Identify potential pitfalls.
- Propose ways to improve performance – e.g., feature selection, regularization, increasing model complexity. Conduct at least **2 more** experiments