

Tabular Data Science – Project Proposal Instructions

1. In the project you need to suggest a new method for improving or automating a specific element within the data science pipeline.
2. The data science “elements” can be anywhere in one (or more) of the following domains:
 1. Exploratory Data Analysis: e.g., Data visualizations, fitting to theoretical distributions, distribution distances
 2. Pattern mining: e.g., Association rules mining, interestingness analysis of patterns
 3. Predictive model analytics: evaluation metrics, feature selection & engineering, analysis of model faults and mispredictions
 4. Explainability
 5. Any other element listed in the current syllabus or taught in class
3. Research proposal Structure (Use the same titles as below, English only)
 1. Project’s Goal:
 1. What element in the DS pipeline are you trying to improve?
 2. Why the chosen element needs improvement?
 3. What are the desired results?
 2. Your solution:
 1. List one or two initial ideas for a solution — this should be very broad and only describe what are you planning to investigate.
 3. Related work:
 1. List 2-3 related works - preferably academic papers but also GitHub repositories or Medium posts are accepted.
 4. Experiments plan (you should test on at least 4 different datasets):
 1. How do you plan to test the effectiveness of your solution?
4. Proposal Submission (by January 4th 2022) :
 1. Send the proposal by email **as a Google Doc**
 2. Don’t forget to add **comments permissions** to the Google Doc.
 3. I may write comments for you to address and notify you.
 4. Check the comments, fix accordingly, and notify me that you’ve fixed them
 5. Once the proposal is approved, register your team and project details in the following [Google Sheet](#)