CMPS 262

Fall 24-25

Capstone-2

In this last milestone, you will be exploring potential datasets to use within the application area you chose from the below:

* Algorithmic Warfare
* Climate and Conflict
* Competitive Intelligence
* Human Resources
* Sustainability and ESG

*Instructions:*

1. Choosing the dataset:

* Consult with online sources (10 pts):
  + open repositories such as (examples but you are not limited to those):
    - <https://climate-conflict.org/www>
    - <https://www.madronavl.com/launchable/public-data-sources-real-estate>
    - <https://www.aihr.com/blog/hr-data-sets-people-analytics/>
    - <https://www.strausscenter.org/ccaps-research-areas/armed-conflict/database/>
  + Datasets described in research papers or books you have identified in Capstone-1
* Format (the more the better):
  + Structured Data in csv files
  + Satellite Images
  + Social Media Data
* Provide your dataset(s) through a link on Google Drive.
* Provide a brief overview of the subject matter of the dataset. What is this dataset about? (10 pts)
* Explain why you chose this dataset(s) in relevance to your application area. (10 pts)
* What are the entities (rows or others) in this dataset(s)? How many are there in this dataset(s)? (10 pts)
* How many attributes (features, variables, predictors) are there in this dataset(s)? What does each one of them designate? (provide a codebook) (10 pts)

1. Describing future analyses to perform later:

* Who are the stakeholders who would be interested in your analyses? Explain. (10 pts)
* What kind of intelligent analyses do you plan on implementing on the chosen datasets? Explain, for example, by choosing one or more from the below (the more, the better): (10 pts)
  + Graph analytics through knowledge graphs
  + Regression (predictive)
  + Classification (predictive)
  + Data exploration for actionable intelligence
* Describe the technologies you are going to use for implementing your solution (programming languages, database management, any scraping needed, machine learning models you would be using, graphical functions you will incorporate, etc.) (10 pts)
* How do you convince potential stakeholders in the value of the intelligent analyses proposed? Explain. (10 pts)
* What are cost-benefit scenarios they need to take into consideration? Explain. (10 pts)
* What could be some challenges (cultural, operational, infra-structure, readiness, etc.) that could pose bottlenecks against the deployment of such a solution in real life? Explain by referring to existing studies if possible. Provide links to such studies if found. (10 pts)

1. Presentation: Prepare a PowerPoint presentation using five slides only to be presented after the deadline has passed.

*Deadline: Monday, October 28, 2024*