Assignment 11

Due on 2020-04-10, 23:59:59

Topic: The use of synthetic or real-world datasets in research.

1. Compare the following two datasets, chosen from the list of available datasets:
   - Datasets A and B
   - Datasets C and D
   - Datasets E and F

   Discuss the characteristics of each dataset, such as size, format, and relevance to the research question.

2. Describe the preprocessing steps required to make each dataset suitable for analysis. Include any features selection, data cleaning, or normalization steps.

3. Analyze the results obtained from analyzing datasets A and B, and discuss the insights gained.

4. Propose a potential application for each of the datasets mentioned above.

5. Write a brief summary of the importance of using synthetic or real-world datasets in research.

6. Choose one dataset from the list and conduct a basic analysis using Python. Include code snippets and brief explanations of your analysis.

7. Discuss the implications of using real-world datasets compared to synthetic datasets in terms of realism and generalizability.

8. Reflect on the ethical considerations involved in using either type of dataset in research.

9. Propose improvements to the datasets mentioned that would enhance their utility for research purposes.

10. Develop a plan for acquiring additional datasets if needed for future research.

11. In your final paragraph, summarize your thoughts on the importance of using datasets in research and your personal experience with the process.

Note: This assignment is worth 20 points.