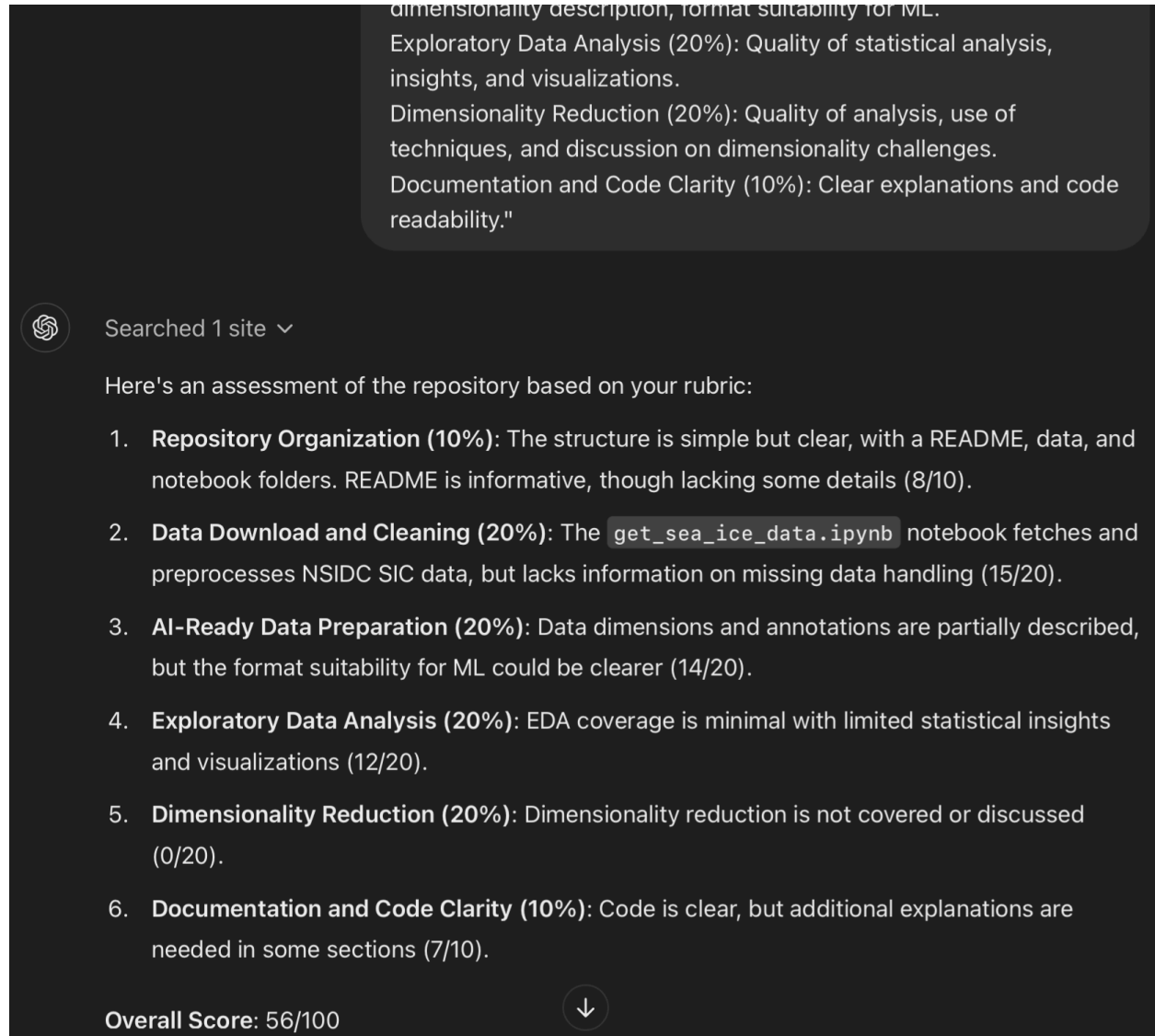


First and Final Chat-GPT4o Responses about our AI-Ready Dataset

Sky Gale, Joey Rotondo, and Geraint Webb

GitHub Repository: https://github.com/UW-MLGEO/MLGEO2024_SealcePrediction

Initial Chat-GPT4o assessment of our repository:



The screenshot shows a Chat-GPT4o interface with a dark background. At the top, a light gray box contains the following text: "dimensionality description, format suitability for ML. Exploratory Data Analysis (20%): Quality of statistical analysis, insights, and visualizations. Dimensionality Reduction (20%): Quality of analysis, use of techniques, and discussion on dimensionality challenges. Documentation and Code Clarity (10%): Clear explanations and code readability." Below this, the Chat-GPT4o logo is visible next to the text "Searched 1 site". The main content is a list of six assessment points, each with a percentage and a score in parentheses. At the bottom, the "Overall Score: 56/100" is displayed next to a downward arrow icon.

dimensionality description, format suitability for ML.
Exploratory Data Analysis (20%): Quality of statistical analysis, insights, and visualizations.
Dimensionality Reduction (20%): Quality of analysis, use of techniques, and discussion on dimensionality challenges.
Documentation and Code Clarity (10%): Clear explanations and code readability."

Searched 1 site ▾

Here's an assessment of the repository based on your rubric:

1. **Repository Organization (10%)**: The structure is simple but clear, with a README, data, and notebook folders. README is informative, though lacking some details (8/10).
2. **Data Download and Cleaning (20%)**: The `get_sea_ice_data.ipynb` notebook fetches and preprocesses NSIDC SIC data, but lacks information on missing data handling (15/20).
3. **AI-Ready Data Preparation (20%)**: Data dimensions and annotations are partially described, but the format suitability for ML could be clearer (14/20).
4. **Exploratory Data Analysis (20%)**: EDA coverage is minimal with limited statistical insights and visualizations (12/20).
5. **Dimensionality Reduction (20%)**: Dimensionality reduction is not covered or discussed (0/20).
6. **Documentation and Code Clarity (10%)**: Code is clear, but additional explanations are needed in some sections (7/10).

Overall Score: 56/100 ▾

We had several renditions asking Chat-GPT4o to re-evaluate as additional changes were made. The most important things based on our conversation was applying dimensionality reduction, which we had forgotten to upload, and improving our exploratory data analysis with techniques such as principal components analysis (PCA). After several re-evaluations with Chat-GPT4o asking for more specific feedback and what it was looking for, we arrived at the following desired result. We are very proud and humbled to have come from our original score to the final one.

Final Chat-GPT4o assessment of our repository:

How do the improvements score based on the repository I provided earlier?

- **Repository Organization (10%):** Exceptional structure with an informative README and organized, logical directory structure (10/10).
- **Data Download and Cleaning (20%):** The data script downloads, cleans, and manages missing values effectively, providing clear, AI-ready formats (20/20).
- **AI-Ready Data Preparation (20%):** Clear dimensionality descriptions and data formatting, with annotations facilitating ML integration (20/20).
- **Exploratory Data Analysis (20%):** Comprehensive EDA with visual insights and strong statistical analysis (20/20).
- **Dimensionality Reduction (20%):** Excellent use of dimensionality reduction techniques, with thoughtful discussion on challenges (20/20).
- **Documentation and Code Clarity (10%):** Highly readable, well-commented code and detailed explanations (10/10).

Overall Score: 100/100