Innovation Lab – DataWaves

Data Science for the Maritime Economy and Environmental Protection

Guidelines for Student Projects

To ensure student projects are feasible within a single academic semester and provide both educational and practical value, the following criteria should be met:

Project Scope and Complexity

- **Ralistic timeframe:** The project should be achievable within 2–3 months, assuming a team of 3–5 students working 4–8 hours per week each (approx. 50–100 hours of total work per student).
- **Clearly defined problem:** The research question should be specific, well-formulated, and achievable within the project timeframe, with measurable outcomes.
- **Appropriate complexity:** The project should not require specialized expertise beyond the standard curriculum of a data science or related degree program.
- **Solution evaluation (benchmark):** A set of criteria, test data, and metrics must be defined to objectively evaluate the quality, performance, and practical value of the proposed solution..

Data Requirements

- **Data availability:** All required data must be provided at the start of the project or within the first few weeks. A data sample should be prepared and shared **during the project proposal phase**.
- **Dataset size:** The dataset should be large enough to support meaningful analysis but small enough to avoid the need for advanced infrastructure (optimal size: 1,000–1,000,000 records or approx. 1MB–10GB).
- **Data quality:** Data should be relatively clean or require only basic preprocessing that students can reasonably perform themselves.
- **Data format:** Preferred formats include CSV, Excel, JSON — easy to process with standard tools — or standard image formats (for computer vision tasks).

Focus on Data Science and Data Analysis

- **Emphasis on analysis:** Projects should focus on data exploration, visualization, modeling, and interpretation.
- **Analytical methods:** Projects may involve statistical analysis, machine learning, predictive analytics, segmentation, pattern recognition, or data visualization.
- **Tools & technologies:** Students are expected to use tools such as Python (with pandas, scikit-learn, matplotlib), R, or big data platforms (e.g., Apache).
- **Expected outcomes:** Final deliverables may include analytical reports, predictive models, source code, and data-driven business recommendations.

Examples of Suitable Topics

- Anomaly detection in operational data
- Process optimization based on historical data
- Optimization of shipping routes with regard to CO_2 emissions
- Forecasting port water pollution levels
- Analyzing the impact of weather conditions on maritime safety
- Predictive maintenance and planning for machinery servicing
- Detecting abnormal vessel behavior / compliance monitoring (fraud detection)
- Route analysis and fleet-wide optimization
- Computer vision for inspections (automated detection of corrosion, leaks, or damage via camera/drone images)
- Pattern detection: fuel consumption vs weather patterns; voyage segmentation/classification; detection of event sequences leading to mechanical failure

Examples of Unsuitable Topics

- Projects requiring the development of complex web or mobile applications
- Tasks that depend on data collection over extended periods (more than 2–3 weeks)
- Projects with no access to concrete datasets
- Projects requiring deep domain knowledge not accessible to students