Qeios

Peer Review

Review of: "Data Integrity vs. Inference Accuracy in Large AIS Datasets"

Rodrigo Pérez Fernández[1]

1. Universidad Politécnica de Madrid, Spain

The paper addresses a crucial topic that is fundamental to maritime safety, traffic management, and environmental monitoring.

It highlights key challenges related to data errors, anomalies, and integrity issues in AIS datasets, making it highly relevant to both academic and operational maritime communities.

The introduction provides a detailed and well-referenced discussion of previous work on AIS data integrity, error detection, and anomaly identification.

References to various machine learning techniques (e.g., Isolation Forest), stochastic methods, and logic-based frameworks indicate a strong theoretical foundation.

The study clearly outlines its scope by focusing on tankers, which is a well-motivated choice due to their operational variability and importance in maritime traffic.

The discussion on outliers in speed data and anomalies in the "Moored" status provides compelling evidence of integrity issues in AIS datasets.

The conclusions effectively highlight the necessity of error detection and correction to improve AIS-based decision-making.

The work has direct implications for the future of autonomous vessel navigation and maritime traffic monitoring.

The abstract is informative, but it could benefit from clearer wording and a stronger statement on the novel contributions of the paper.

For example, specifying which error detection methods were tested and summarizing quantitative improvements in classification accuracy would enhance clarity.

While the literature review is extensive, the main research question is somewhat diluted within the numerous references.

The study mentions that AIS data was purchased from S&P Global (IHS Markit), but details on data

doi.org/10.32388/TPR6SW

preprocessing steps, cleaning methods, and feature selection are lacking.

How were outliers detected and corrected? Were specific machine learning models applied to test the impact on classification accuracy?

While anomalies such as ships moving at certain knots in "Moored" status are discussed, the study could quantify the impact of data integrity improvements on inference accuracy.

The conclusion effectively summarizes findings but could be expanded to include:

How specific integrity improvements translated into improved accuracy.

Future research directions, e.g., the role of AI-driven anomaly detection in AIS data processing.

Potential real-world applications of the findings for coastal authorities, autonomous shipping, and maritime law enforcement.

Declarations

Potential competing interests: No potential competing interests to declare.