Qeios

Peer Review

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

Rossi Passarella[1]

1. Universitas Sriwijaya, Indonesia

Abstract: The problem statement is clear, the focus is on the right things, and the solution and impact are discussed. However, there are some issues, such as not being clear about the type of classification task that was considered, the error detection and correction method that was used, or the results that could be measured. What kind of classification? How much improvement has been made? In addition, the contribution is illegible, and there is no mention of specific datasets such as size or resources.

Introduction: In the introduction, especially in the last paragraph of this section, I conclude as follows: This paragraph, while outlining the general context of AIS data integrity and its importance for maritime domain awareness, suffers from several weaknesses that hinder its effectiveness. The language used is often vague and lacks precision, employing generic phrases like "collectively aim," "improve data quality," and "mitigate risks" without specifying how these goals are achieved or the nature of the risks. The connection between the stated aims and the specific database from S&P Global (IHS Markit) is weak, creating a disjointed flow of ideas. Furthermore, the phrasing "data is (or, at least, should be) sourced" is awkward and unprofessional, casting doubt on the data's origin. The redundant use of "data quality" and "data integrity" without clear distinction, coupled with the passive construction, "In this study we proposed," further weakens the writing. Most significantly, the paragraph omits any mention of the specific methodology employed for data integrity analysis, a critical omission for a scientific publication. Finally, the grammatical error "Chapter 2 consists" and the generic statement about the article concluding with conclusions add no value. In its current state, the paragraph lacks the clarity, precision, and essential details required for a strong academic contribution.

doi.org/10.32388/C5PNQS

Data analysis for integrity: In Section 2, when describing the data, it would be helpful to provide the dimensions of the data rows purchased from IHS Markit. For instance: 'The data rows used in this study have dimensions of 3256x240 (rows x columns). This structure, with its [mention key aspects of the dimensions, e.g., large number of rows and diverse columns], facilitates a comprehensive analysis of various parameters and allows us to draw meaningful insights from the extensive IHS Markit dataset.

Declarations

Potential competing interests: No potential competing interests to declare.