

Mining imperfect data: dealing with contamination and incomplete records

Pearson, Ronald K., 1952- , author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20443143&lokasi=lokal>

Abstrak

Data mining is concerned with the analysis of databases large enough that various anomalies, including outliers, incomplete data records, and more subtle phenomena such as misalignment errors, are virtually certain to be present. Mining Imperfect Data describes in detail a number of these problems, as well as their sources, their consequences, their detection, and their treatment. Specific strategies for data pretreatment and analytical validation that are broadly applicable are described, making them useful in conjunction with most data mining analysis methods. Examples are presented to illustrate the performance of the pretreatment and validation methods in a variety of situations, both simulation based, where "correct" results are known unambiguously, and real data examples that illustrate typical cases met in practice.