

# STATISTICAL ANALYSIS AND DATA MINING

Original Article

## Measurement of data complexity for classification problems with unbalanced data

Nafees Anwar ✉, Geoff Jones, Siva Ganesh

First published: 02 May 2014

<https://doi.org/10.1002/sam.11228>

Cited by: 3



About



Access

als



ire

### Abstract

We introduce a complexity measure for classification problems that takes account of deterioration in classifier performance as a result of class imbalance. The measure is based on  $k$  - nearest neighbors. We explore the choices of  $k$  and the distance metric through a simulation study, and illustrate the use of our measure, and related data visualization techniques, with real datasets from the literature.

### Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley

