

## Optimizing Communications Overhead in Heterogeneous Distributed Data Streams

**Authors :** Rashi Bhalla, Russel Pears, M. Asif Naeem

**Abstract :** In this 'Information Explosion Era' analyzing data 'a critical commodity' and mining knowledge from vertically distributed data stream incurs huge communication cost. However, an effort to decrease the communication in the distributed environment has an adverse influence on the classification accuracy; therefore, a research challenge lies in maintaining a balance between transmission cost and accuracy. This paper proposes a method based on Bayesian inference to reduce the communication volume in a heterogeneous distributed environment while retaining prediction accuracy. Our experimental evaluation reveals that a significant reduction in communication can be achieved across a diverse range of dataset types.

**Keywords :** big data, bayesian inference, distributed data stream mining, heterogeneous-distributed data

**Conference Title :** ICCSDM 2020 : International Conference on Computer Science and Data Mining

**Conference Location :** Auckland, New Zealand

**Conference Dates :** December 01-02, 2020