



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Seminar

**APPLICATION OF THE VARIABLE
PRECISION ROUGH SETS MODEL
TO ESTIMATE THE OUTLIER
PROBABILITY OF EACH ELEMENT**

Alberto Fernández Oliva
University of Havana

**June 14, 2019 | 12.30 p.m. | Aula Benvenuti
Campus S. Caterina**

Abstract : www.stat.unipd.it/fare-ricerca/seminari

APPLICATION OF THE VARIABLE PRECISION ROUGH SETS MODEL TO ESTIMATE THE OUTLIER PROBABILITY OF EACH ELEMENT

Alberto Fernández Oliva

Full Professor | University of Havana (Cuba)

In a data mining process, outlier detection aims to use the high marginality of these elements to identify them by measuring their degree of deviation from representative patterns, thereby yielding relevant knowledge. Whereas rough sets (RS) theory has been applied to the field of knowledge discovery in databases (KDD) since its formulation in the 1980s; in recent years, outlier detection has been increasingly regarded as a KDD process with its own usefulness. The application of RS theory as a basis to characterise and detect outliers is a novel approach with great theoretical relevance and practical applicability. However, algorithms whose spatial and temporal complexity allows their application to realistic scenarios involving vast amounts of data and requiring very fast responses are difficult to develop. In this seminar we present a theoretical framework based on a generalisation of RS theory, termed the variable precision rough sets model (VPRS), which allows the establishment of a stochastic approach to solving the problem of assessing whether a given element is an outlier within a specific universe of data. An algorithm derived from quasi-linearisation is developed based on this theoretical framework, thus enabling its application to large volumes of data.

The experiments conducted demonstrate the feasibility of the proposed algorithm, whose usefulness is contextualised by comparison to different algorithms analysed in the literature.