Data mining and machine learning to enhance new-particle formation identification and analysis

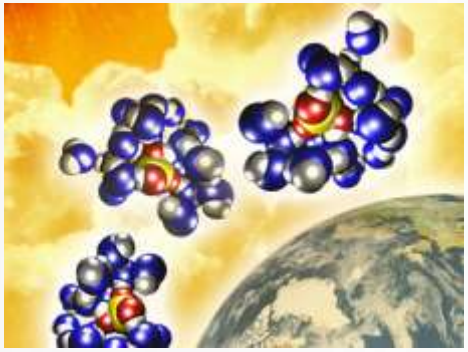
Martha Arbayani Zaidan

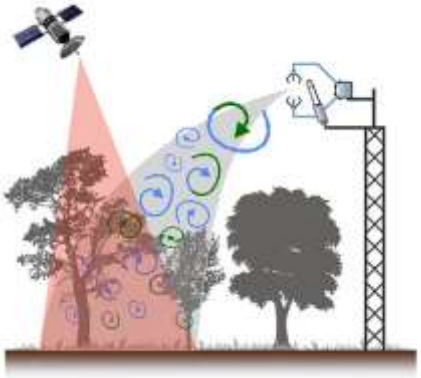
EGU PEEX Special Session
7 May 2020

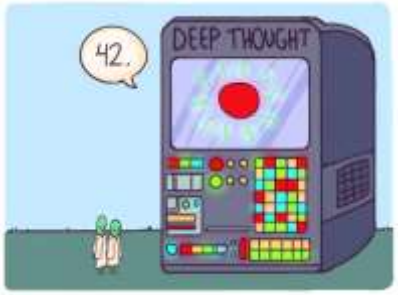
Big atmospheric data

Atmosphere is complex due to the physical and chemical processes involved

Research campaigns, Satellites, Research stations, Experimental chambers, Simulation models, IoT







Raw Data

BIG DATA

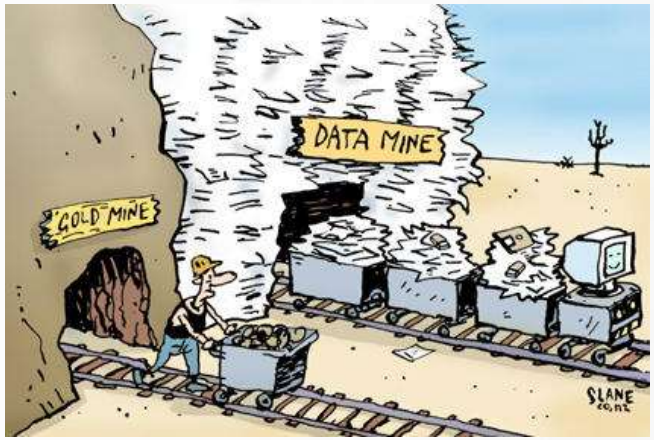
Features Extraction

Data Mining

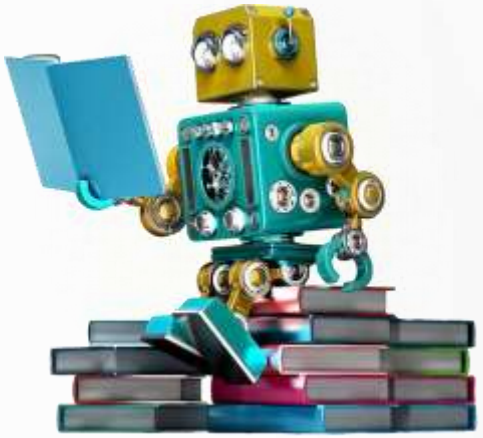
Useful information

Data mining & Machine learning

Data Mining can be defined as the process of analyzing hidden patterns of data to provide useful information.



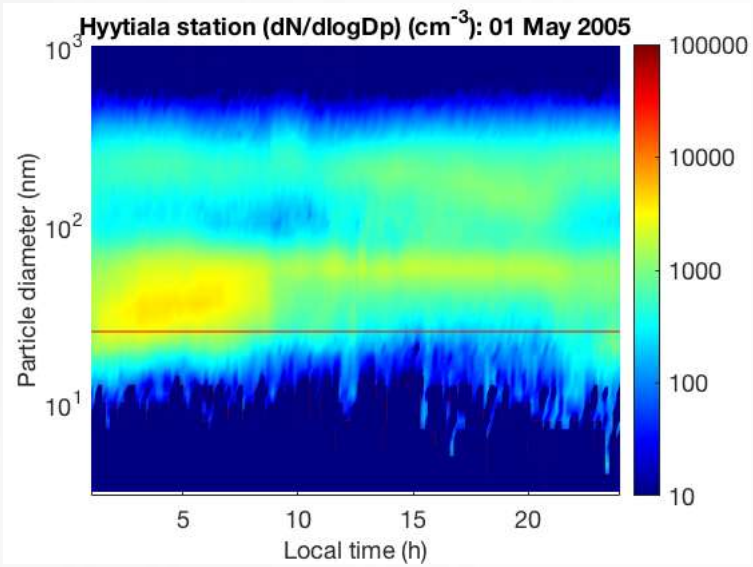
Machine Learning is the science of making a computer (or machine) learn from data without being explicitly programmed, the machine is then subsequently able to perform automatic tasks.



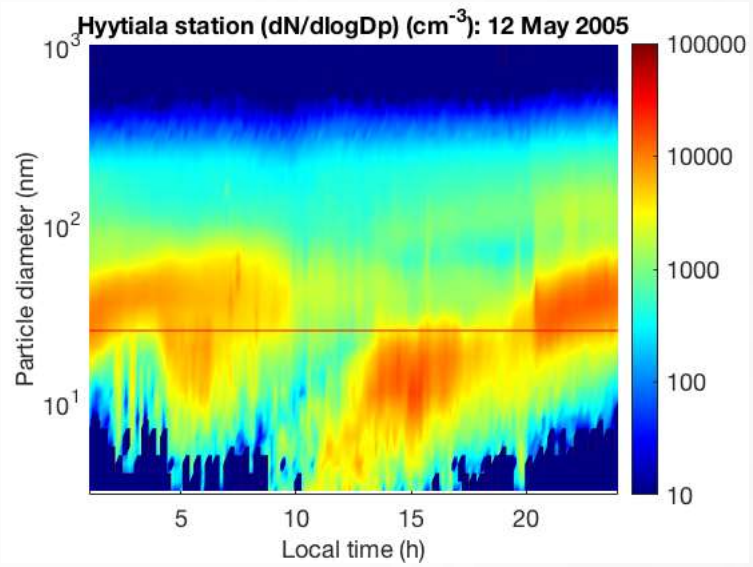
Atmospheric New Particle-Formation (NPF)

NPF is an important source globally of climatically-relevant atmospheric aerosols.

Non-Event day

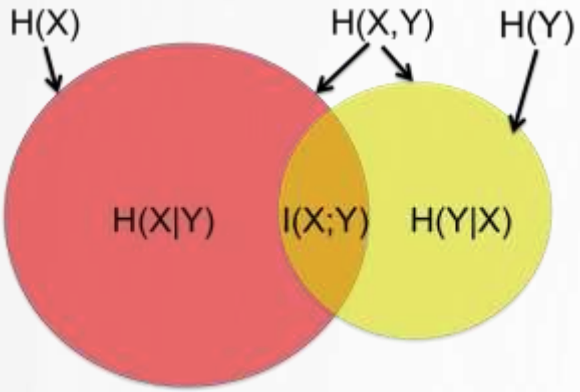


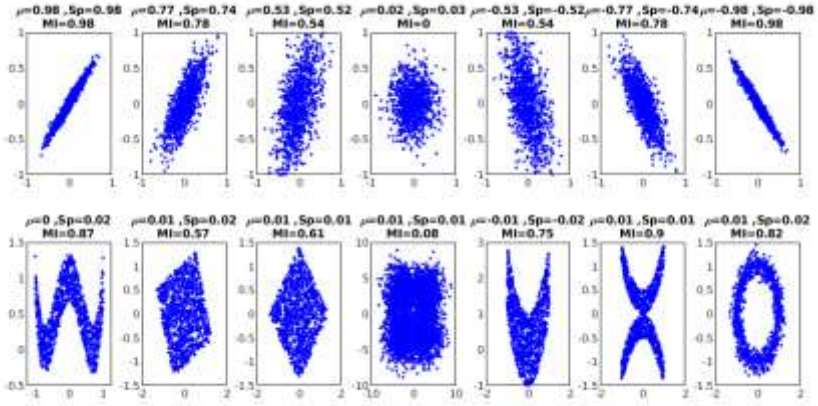
Event day

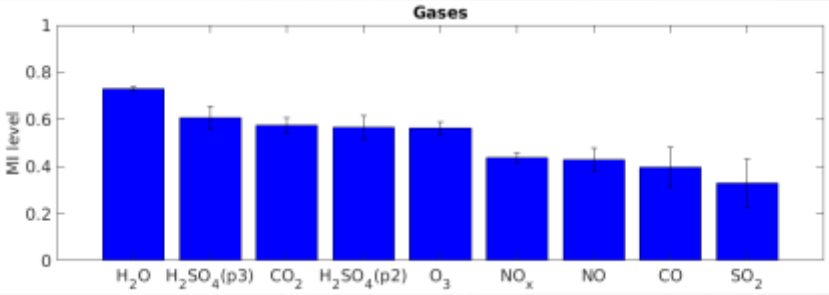


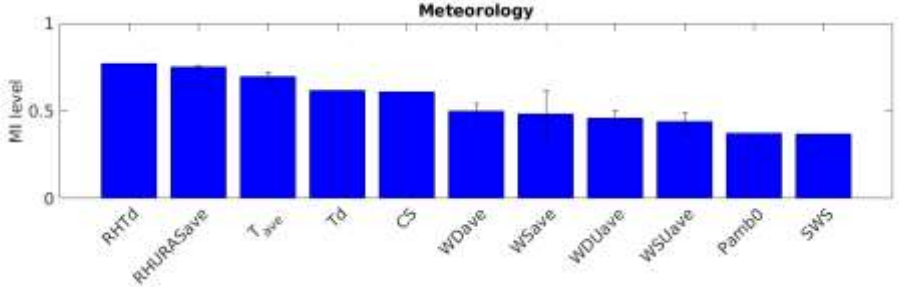
INAR
INSTITUTE FOR ATMOSPHERIC AND EARTH SYSTEM RESEARCH

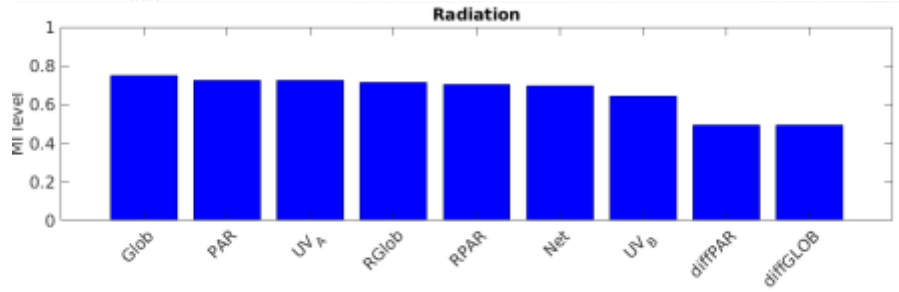
Mutual Information



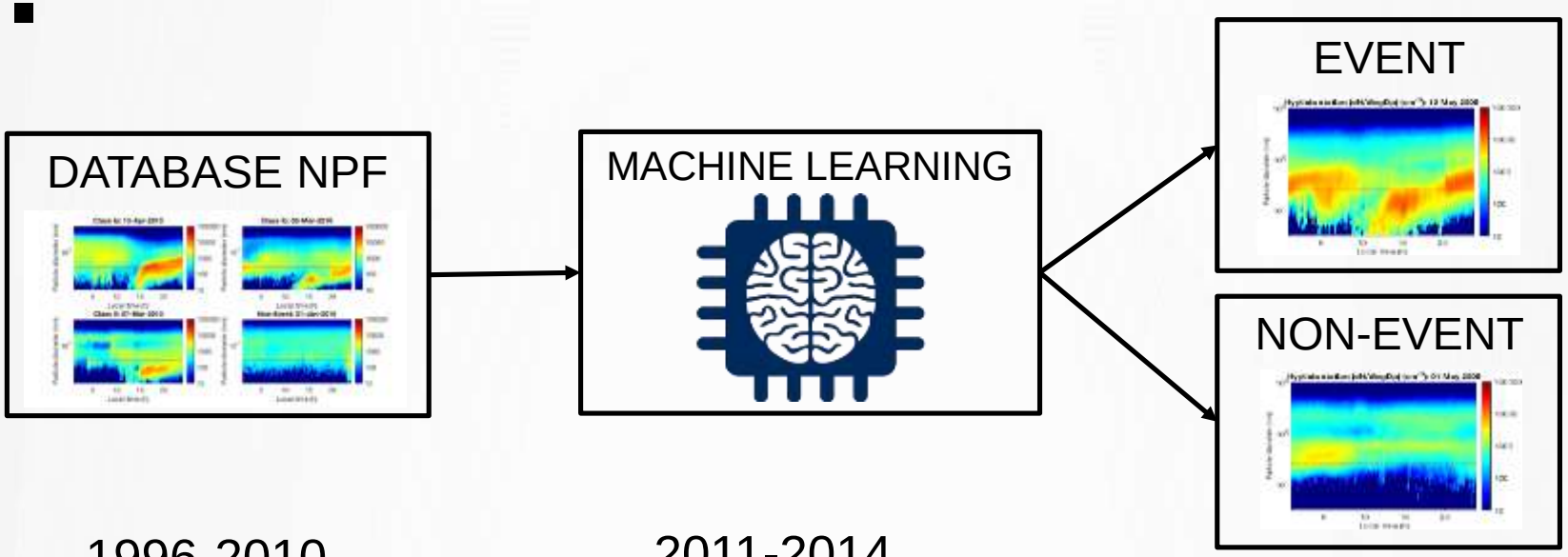








Machine learning classifier



1996-2010

Machine Learning methods	Visualization methods: Event-day	Non-event days	
Event-day	1229 42.2%	38 1.3%	97%
Non-event days	26 0.9%	1621 55.6%	98.4%
	97.9%	97.7%	97.8%

2011-2014

Machine Learning methods	Visualization methods: Event-day	Non- event days	
Event-day	251 31.1%	75 9.3%	77%
Non-event days	59 7.3%	423 52.4%	87.8%
	81.0%	84.9%	83.4%