



International Center for Tropical Agriculture  
*Since 1967 Science to cultivate change*

## Big Data

Marzo, 2019

Juan Camilo Rivera  
[j.c.rivera@cgiar.org](mailto:j.c.rivera@cgiar.org)

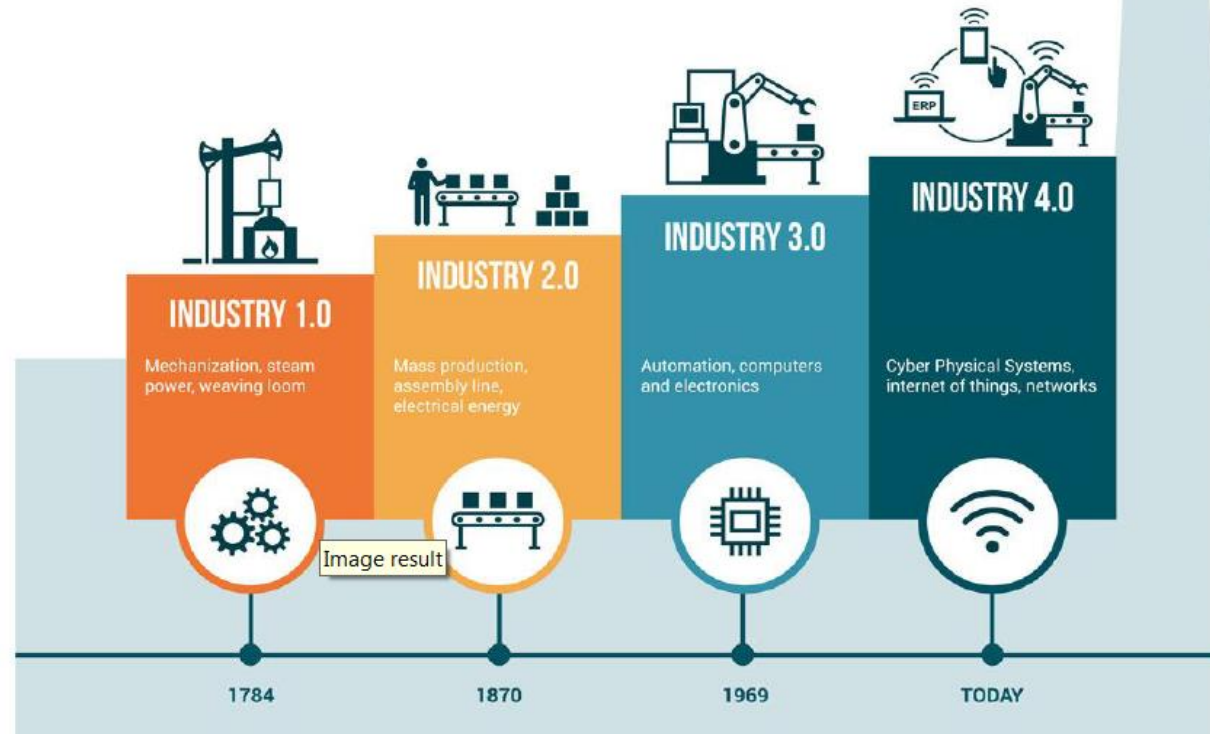
Hugo Andres Dorado  
[h.a.dorado@cgiar.org](mailto:h.a.dorado@cgiar.org)



CGIAR  
CIAT is a CGIAR Research Center

# INDUSTRIAL REVOLUTION

TRANSFORMING INDUSTRIES AND INNOVATION

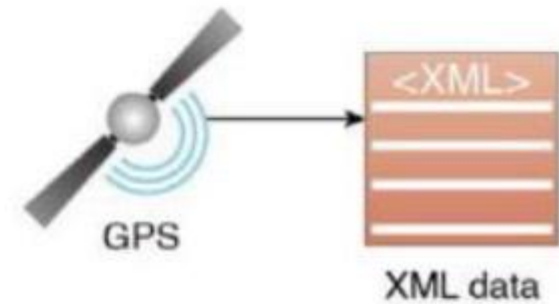
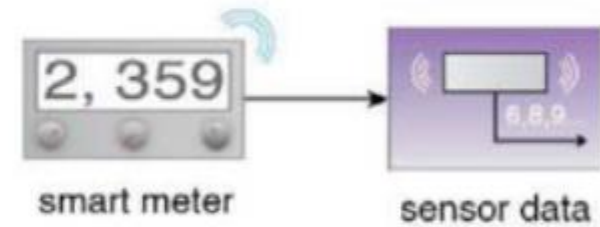
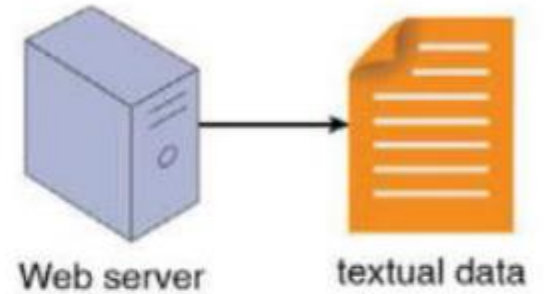
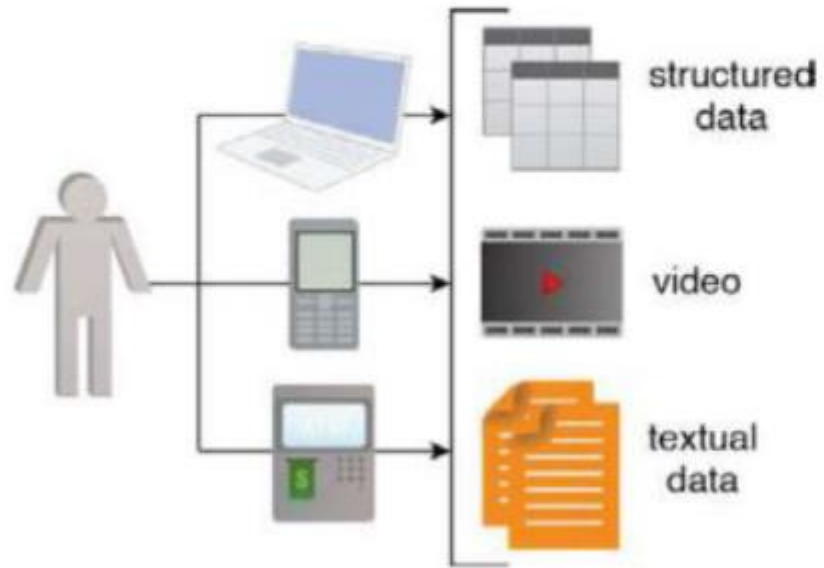


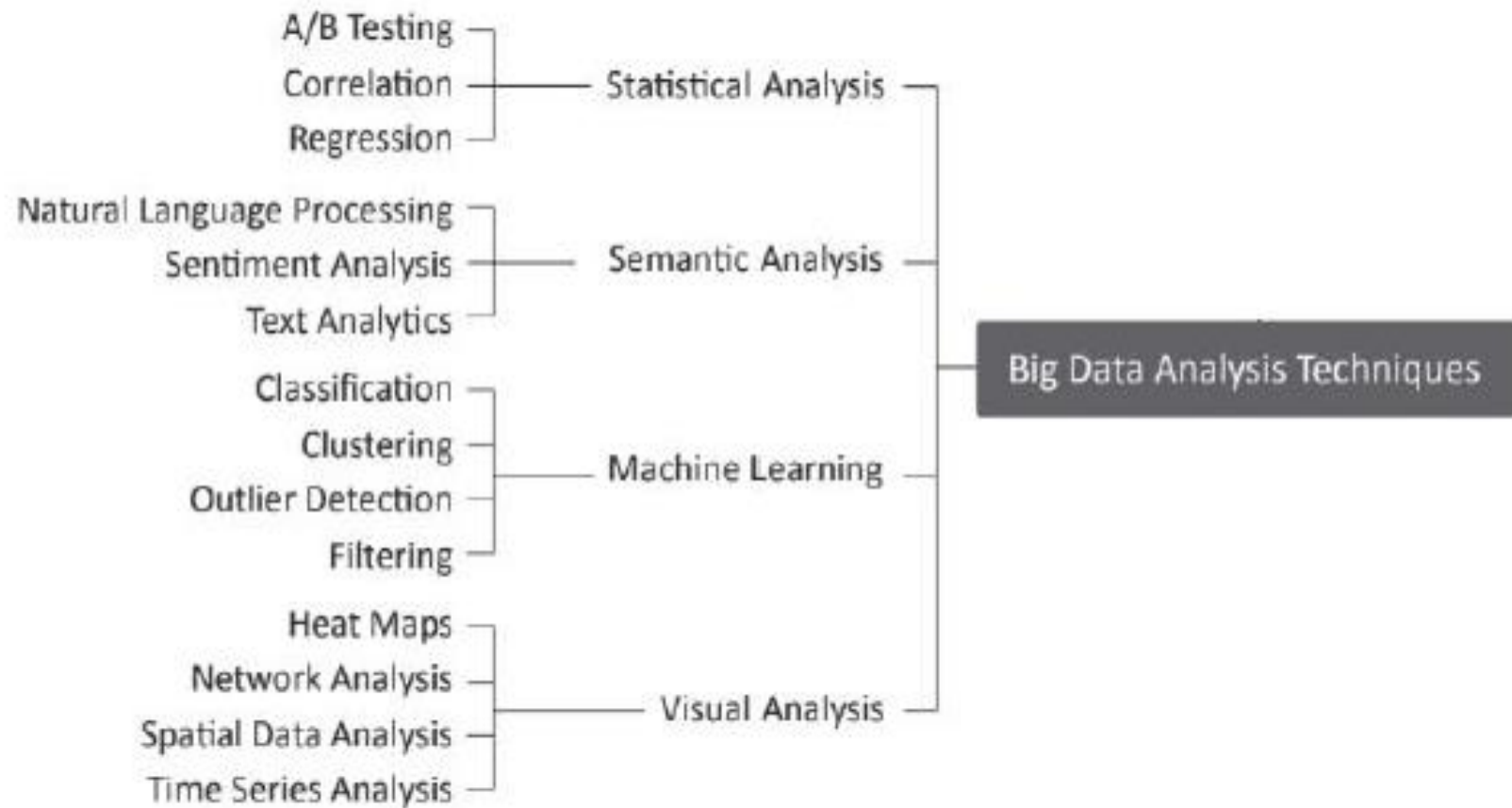




## Big Data Sources

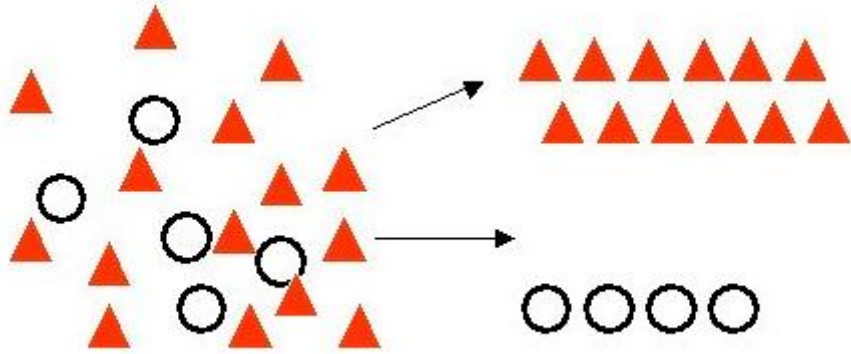
human-generated  
machine-generated



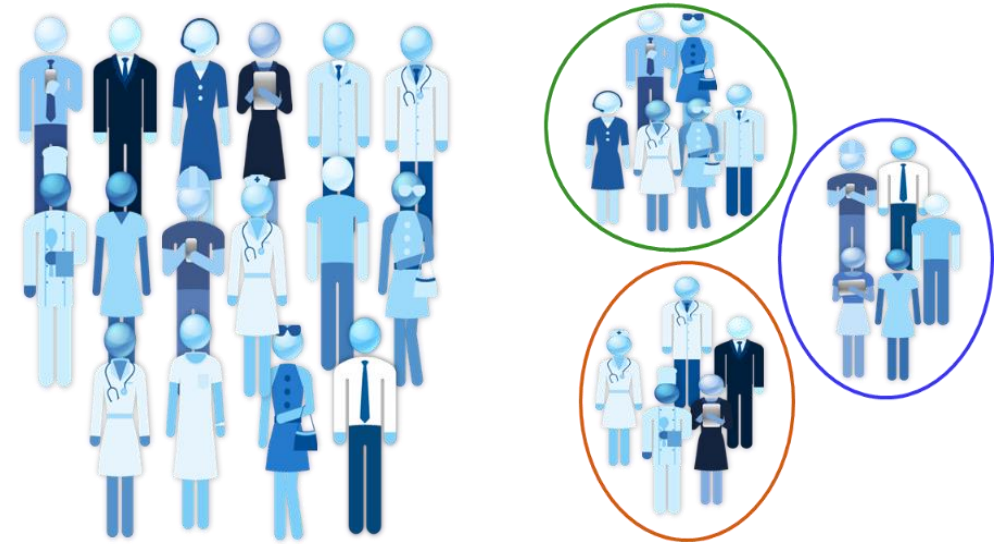


# Tres tareas en minería de datos

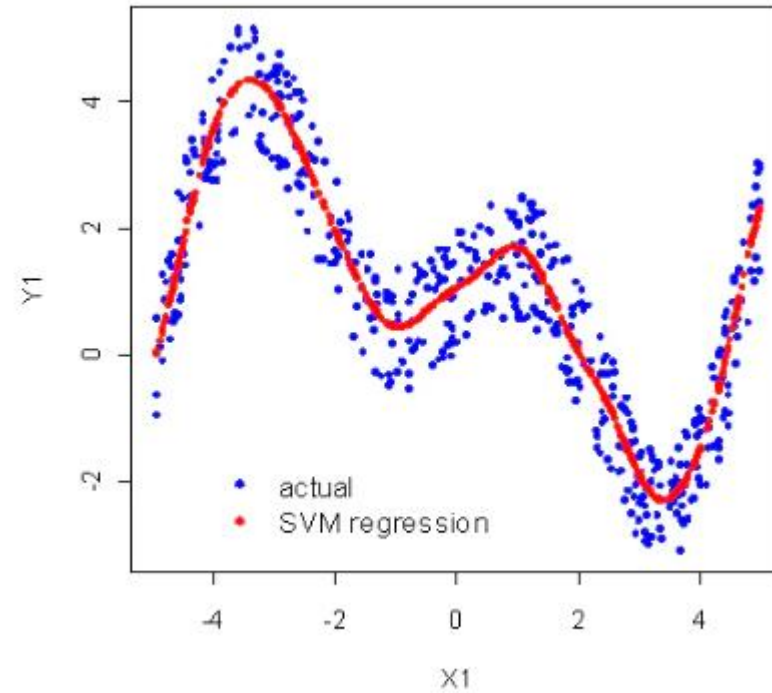
- Clasificación



- Agrupación



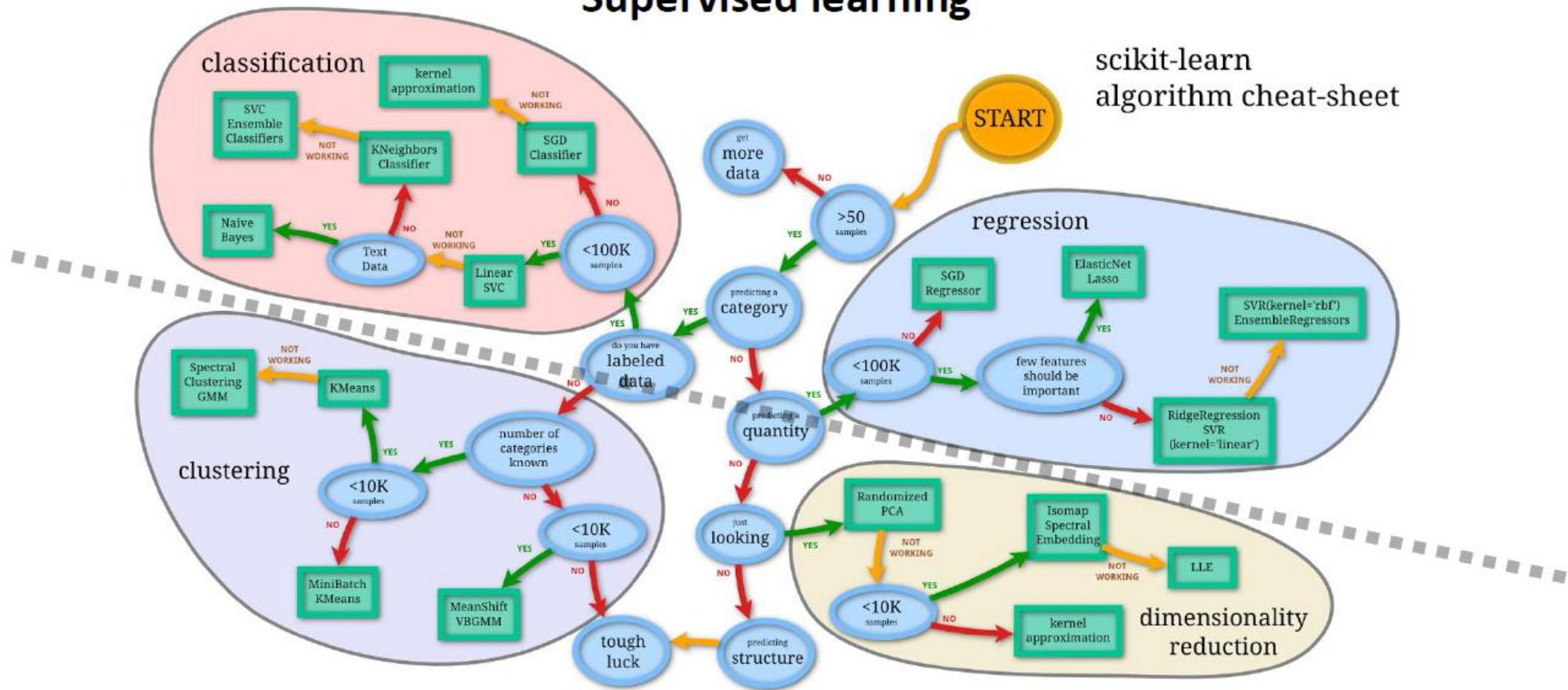
- Regresión



# Machine Learning

## Supervised learning

scikit-learn  
algorithm cheat-sheet



## Unsupervised learning



## TYPES OF PREDICTIVE MODELS AND TECHNIQUES

### LINEAR MODELS



Linear models and linear regression techniques are the most fundamental methods for predictive modeling. These models describe the relationship between two or more independent variables.

### GENERALIZED ADDITIVE MODELS



Generalized additive models (GAM) are a type of *nonparametric* regression, meaning that it does not confirm to any assumptions, as is the case with linear models, which assume linearity.

### LOGISTIC REGRESSION



Logistic regression is used to model the relationship between a categorical dependent variable and one or more response measures, by predicting the probability that a case belongs to a target class.

### SURVIVAL ANALYSIS



Survival analysis is a group of techniques that models the time duration until one or more events happen. It is sometimes referred to as reliability or duration analysis.

### DECISION TREE LEARNING



Decision tree learning is a class of methods where the output is a list of rules that progressively segment the population into smaller, homogeneous groups.

### BAYESIAN METHODS



Bayesian methods are a group of techniques based on Bayes' rule, where the probability of an event is conditional based on different relationships between events.

# Referencias

<http://www.bigdatapatterns.org>

# Thank you!



WE'RE PROUD TO  
HAVE CELEBRATED 50 YEARS  
OF AGRICULTURAL RESEARCH  
FOR DEVELOPMENT

**International Center for Tropical Agriculture - CIAT**

Headquarters and Regional Office  
for South America and the Caribbean

+57 2 445 0000

Km 17 Recta Cali-Palmira  
A.A. 6713, Cali, Colombia

✉ [ciat.cgiar.org](mailto:ciat.cgiar.org)

🌐 [ciat.cgiar.org](http://ciat.cgiar.org)



CIAT is a CGIAR Research Center