

# Selecting electrical billing attributes: big data preprocessing improvements

## Autores

Amelec Vilorio, Jesús García Guilianny, Nataly Orellano Llinás, Hugo Hernandez-P, Ernesto Steffens Sanabria, Omar Bonerge Pineda Lezama.

## Abstract

The attribute selection is a very relevant activity of data preprocessing when discovering knowledge on databases. Its main objective is to eliminate irrelevant and/or redundant attributes to obtain computationally treatable issues, without affecting the quality of the solution. Various techniques are proposed, mainly from two approaches: *wrapper* and *ranking*. This article evaluates a novel approach proposed by Bradley and Mangasarian (Machine learning ICML. Morgan Kaufmann, San Fco, CA, pp. 82–90, 1998 [\[1\]](#)) which uses concave programming for minimizing the classification error and the number of attributes required to perform the task. The technique is evaluated using the electric service billing database in Colombia. The results are compared against traditional techniques for evaluating: attribute reduction, processing time, discovered knowledge size, and solution quality.

## Palabras clave

Electric billing, Concave programming, Data mining, Electric service billing