

# Informe trabajo práctico número uno - Aprendizaje Automático

Omar Ernesto Cabrera Rosero  
Universidad de Buenos Aires  
Email: omarcabrera@udenar.edu.co

Jimmy Mateo Guerrero Restrepo  
Universidad de Buenos Aires  
Email: jimaguere@gmail.com

## Resumen

En este trabajo práctico se analizan las particularidades de la utilización de algoritmos para la generación de árboles de decisión, para la realización de este trabajo se utilizó un conjunto de datos de las pruebas de estado saberpro que se realizan en Colombia a estudiantes universitarios ...

## Keywords

*árboles de decisión, J48, ICFES, saberpro*

## I. INTRODUCCIÓN

Aquí va la intruducción y la tabla de las 31 variables

## II. SOBREAJUSTE Y PODA

texto

En la figura 1 se muestra la grafica el número de hojas en función de la función de poda.

En la figura 2 se muestra la grafica el performance en función de la función de poda.

En la figura 3 se muestra la grafica de la curva ROC para el mejor árbol,

En la figura 4 se muestra la grafica de la curva ROC para el mejor árbol,

En la figura 5 se muestra la grafica de la curva ROC para el mejor árbol,

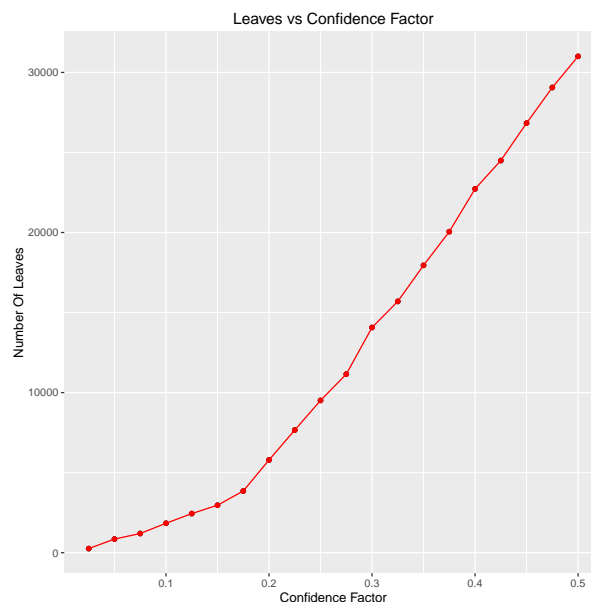


Figura 1. Number of leaves vs Confidence factor

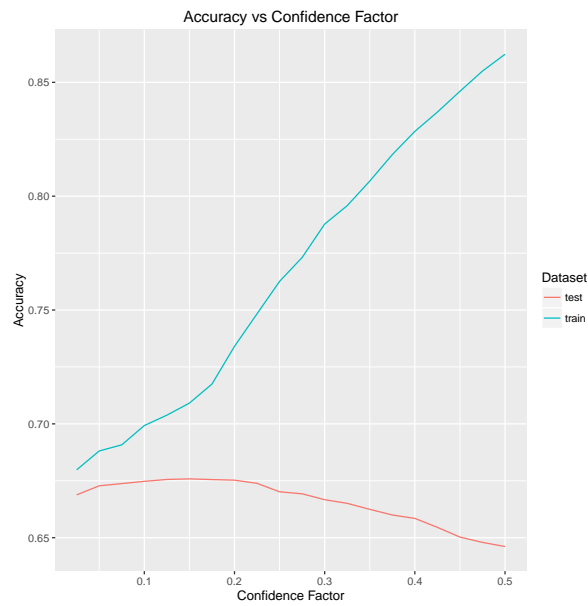


Figura 2. Accuracy vs Confidence factor

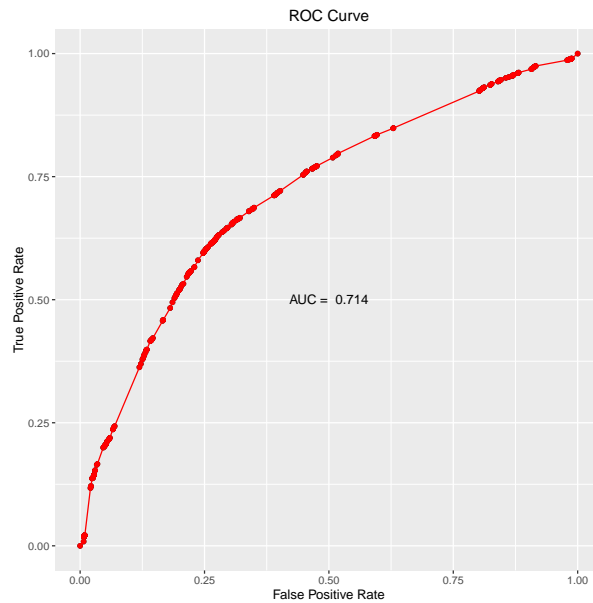


Figura 3. Curva ROC mejor árbol

En la figura 6 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 7 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 8 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 9 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 10 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 11 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 12 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 13 se muestra la grafica de la curva ROC para el mejor árbol,  
 En la figura 14 se muestra la grafica de la curva ROC para el mejor árbol,

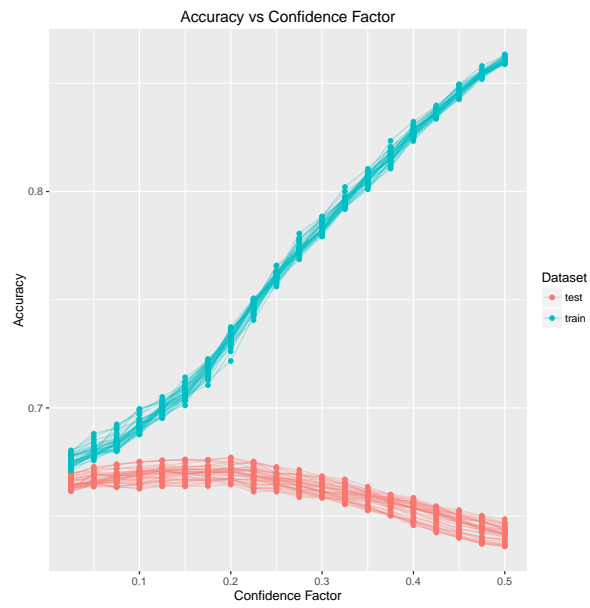


Figura 4. Accuracy vs Confidence factor with missing data

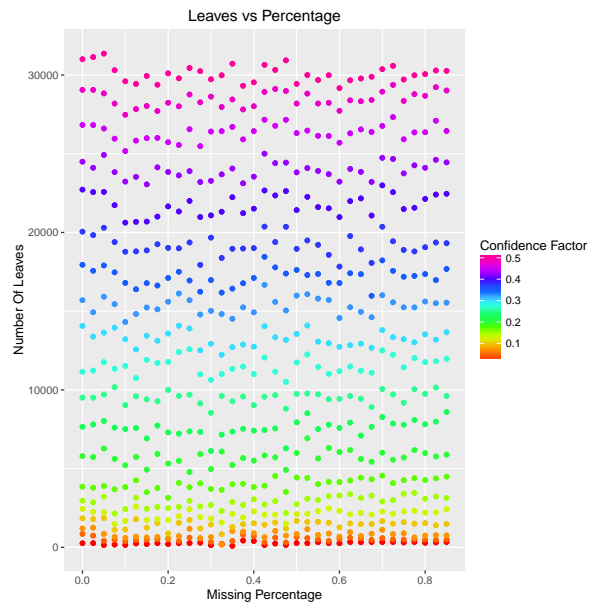


Figura 5. Leaves vs missing percentage

En la figura 15 se muestra la grafica de la curva ROC para el mejor árbol,

En la figura 16 se muestra la grafica de la curva ROC para el mejor árbol,

En la figura 17 se muestra la grafica de la curva ROC para el mejor árbol,

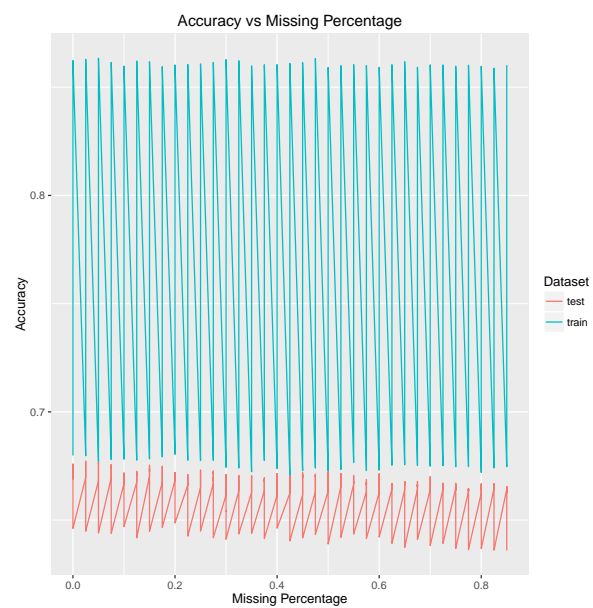


Figura 6. Accuracy vs missing percentage

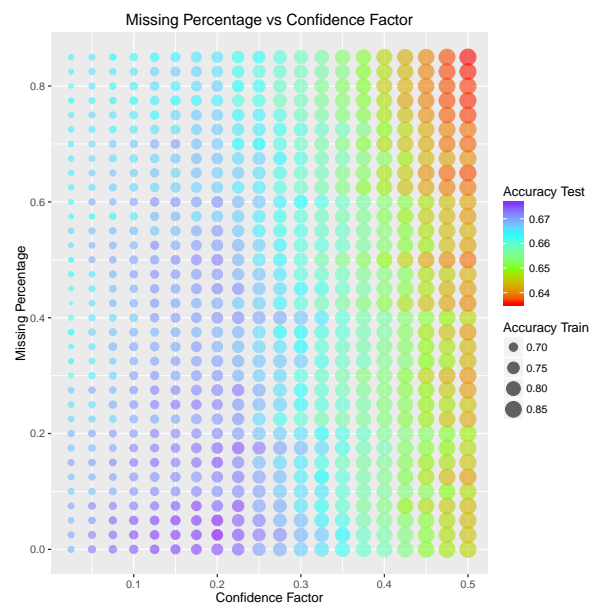


Figura 7. Missing percentage vs Confidence factor

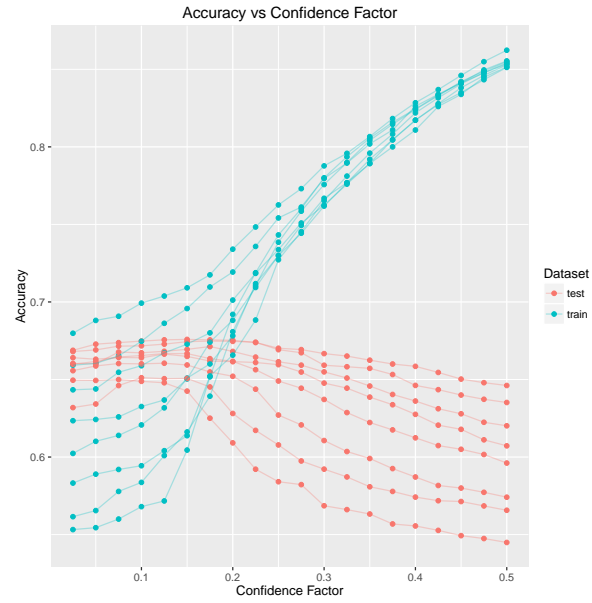


Figura 8. Accuracy vs Confidence factor with noise data

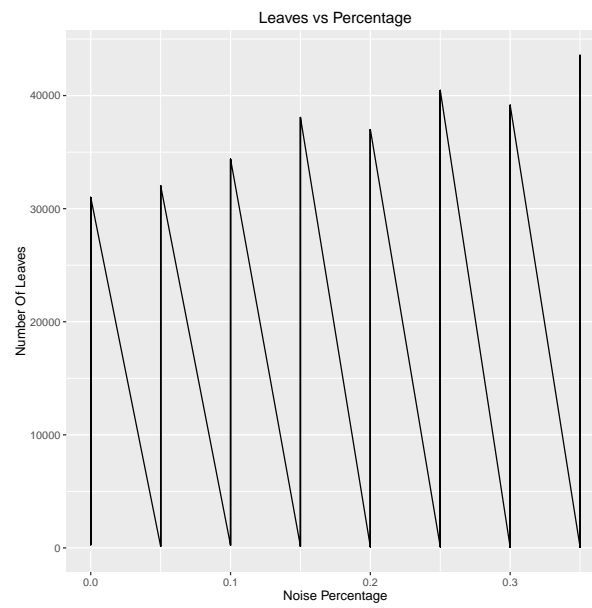


Figura 9. Leaves vs noice percentage

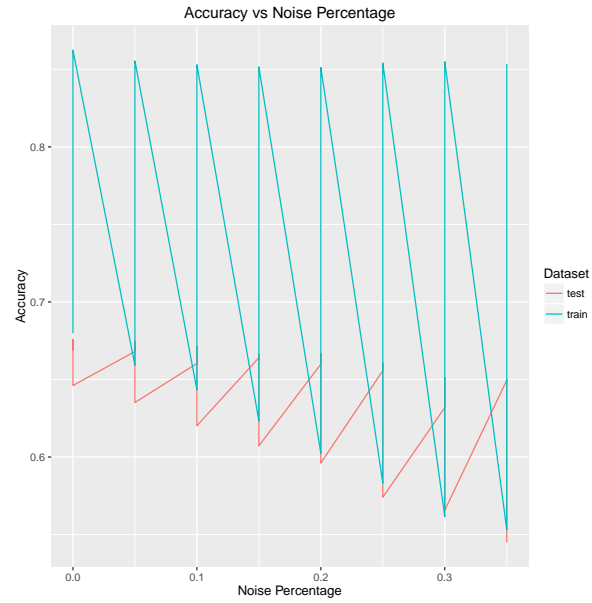


Figura 10. Accuracy vs noise percentage

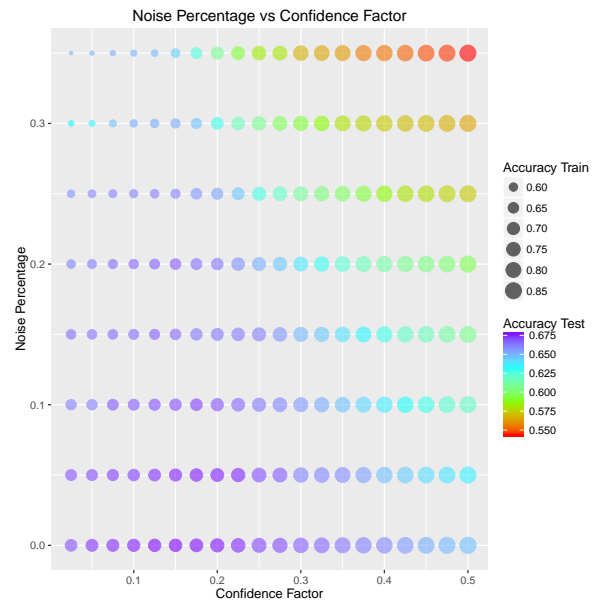


Figura 11. Noise percentage vs Confidence factor

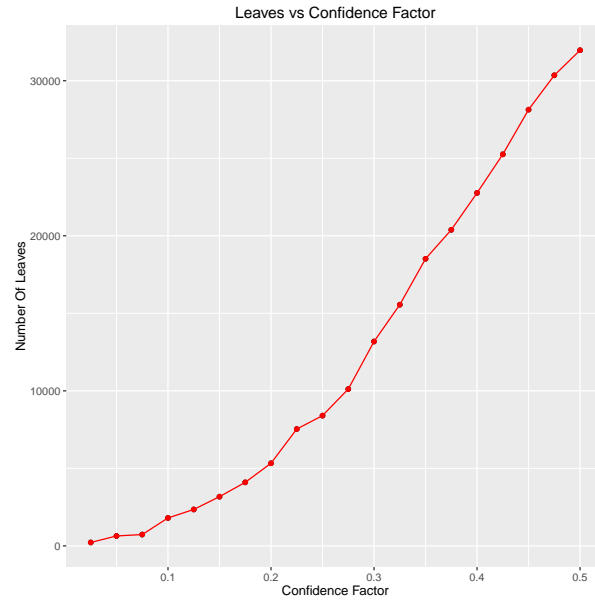


Figura 12. Number of leaves vs Confidence factor with supervised discretize

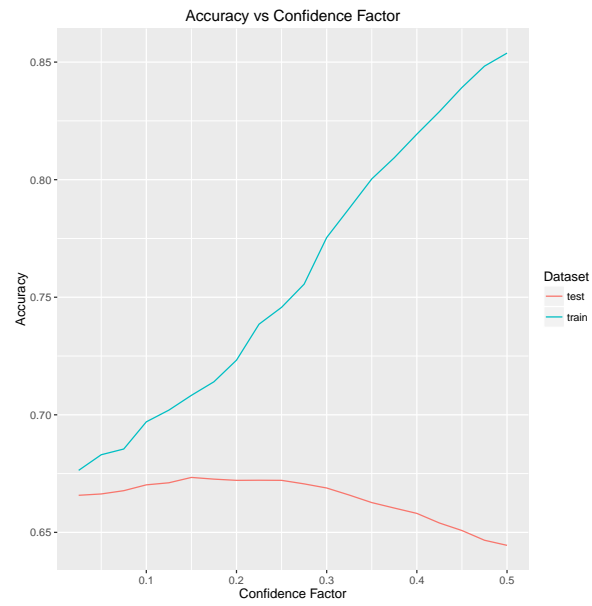


Figura 13. Accuracy vs Confidence factor with supervised discretized

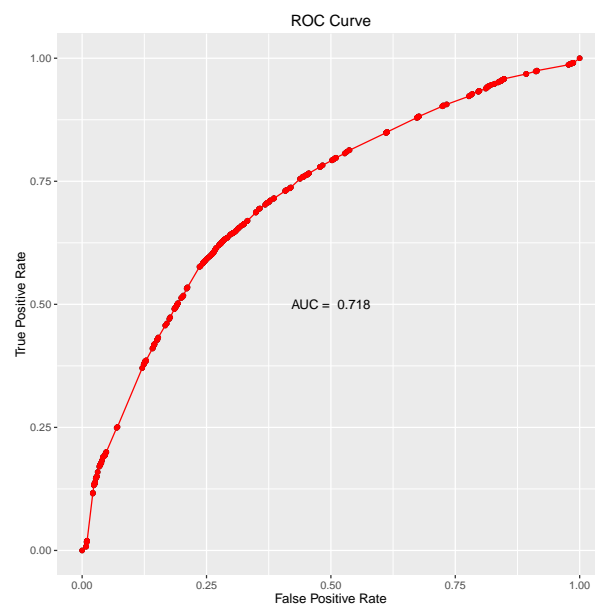


Figura 14. Curva ROC mejor árbol with supervised discretized

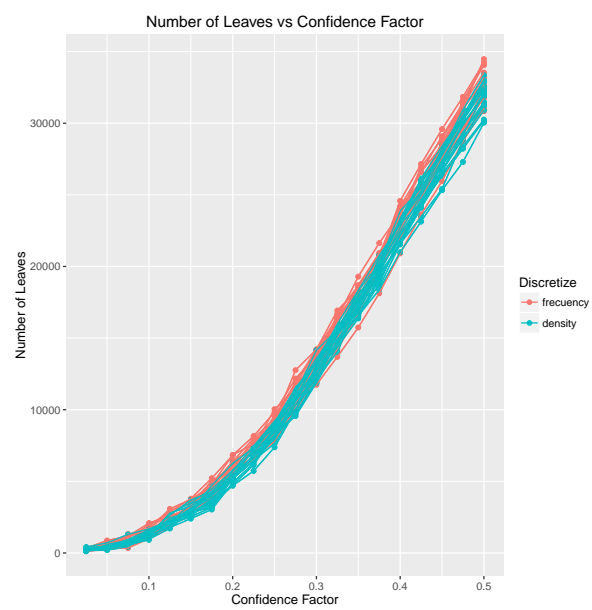


Figura 15. Leaves vs missing percentage with discretize



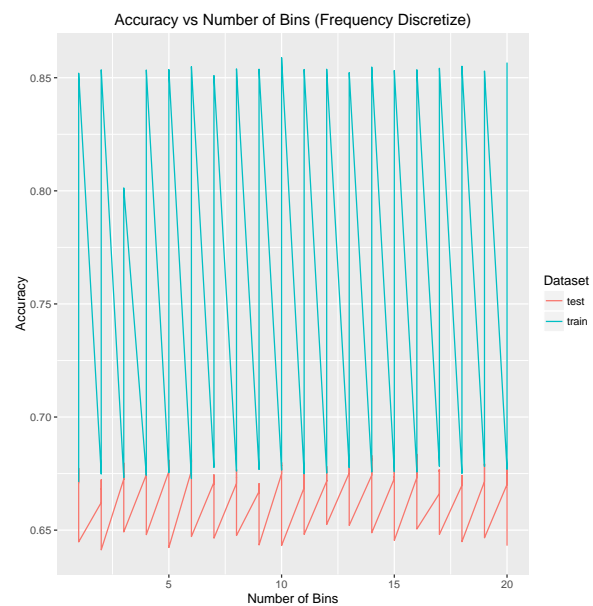


Figura 16. Accuracy vs Number of bins (Frequency discretize)

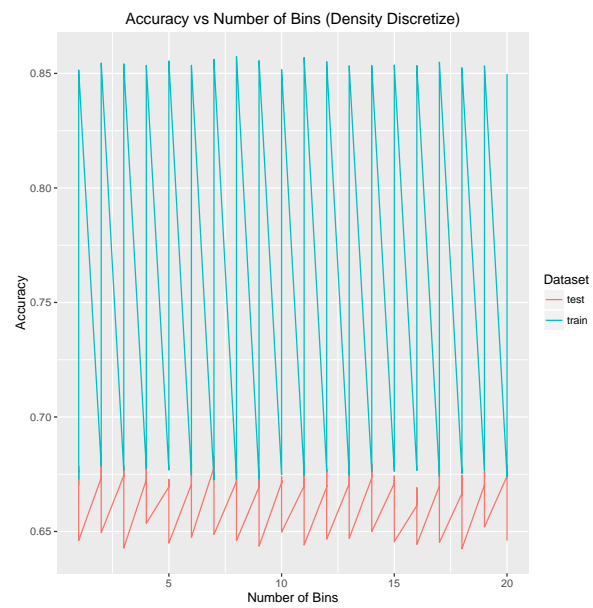


Figura 17. Accuracy vs Number of bins (Density discretize)