
**CLASIFICACIÓN BINARIA MEDIANTE *ENSEMBLES* PARA CLASES
DESBALANCEADAS: UNA APLICACIÓN AL PROBLEMA DE
PREDICCIÓN DE FUGA DE CLIENTES****CLAUDIO ALFONSO MONTECINOS VALENZUELA
MAGÍSTER EN GESTIÓN DE OPERACIONES****RESUMEN**

En el presente trabajo de Tesis se propone una metodología para la construcción de un modelo de clasificación binaria en el contexto de desbalance y traslape de clases, con el objetivo de generar un modelo con un mayor poder predictivo. El modelo está basado en la combinación de métodos de dos enfoques de clasificación: el enfoque discriminante y el basado en reconocimiento o clasificación de una clase. A través de esta estrategia se espera inducir clasificadores individuales capaces de rescatar de mejor manera diversas regiones del espacio de datos y de esta manera combinar sus decisiones en un *ensemble* de métodos. Al respecto, dos estrategias de combinación de decisiones son utilizadas: reglas fijas de combinación y el método *Stacking*. La aplicación de la metodología sobre bases de datos artificiales muestra la capacidad de los *ensembles* para incrementar el rendimiento de clasificación. Al respecto, los beneficios generados son mayores en aquellas bases de datos más complejas, en las cuales los métodos individuales no generan resultados satisfactorios, pero tienden a ser más diversos. Además, la metodología es aplicada a un problema real de predicción de fugas de clientes. En este contexto, se confirma la capacidad de los *ensembles* para generar predicciones más asertivas, logrando un 5.4% de incremento por sobre el máximo alcanzado por un clasificador individual, considerando como medida de rendimiento la media geométrica de las tasas de acierto individuales sobre cada clase. Se destaca que este beneficio es originado principalmente por un incremento en la tasa de acierto sobre la clase "Fuga" (minoritaria). Los resultados obtenidos en la etapa de experimentación, así como en la aplicación real, sugieren que el método *Stacking* es el más adecuado para la combinación de clasificadores cuando es inducido un meta-clasificador robusto como SVM con función de kernel RBF.