

## RESUMEN

Autor [Pariona Huarhuachi, J.C.](#)  
Autor [Universidad Nacional Agraria La Molina, Lima \(Peru\).](#)  
corporativo [Facultad de Economía y Planificación](#)  
Título **Clasificación de fuga de clientes en una entidad financiera utilizando el algoritmo SMOTE para datos desbalanceados en una regresión logística**  
Impreso Lima : UNALM, 2017

### Copias

Ubicación	Código	Estado
Sala Tesis	<a href="#">E13. P37 - T</a>	EN PROCESO
Descripción	88 p. : 17 fig., 25 cuadros, 43 ref. Inlcuye CD ROM	
Tesis	Tesis (Ing Estadístico e Informático)	
Bibliografía	Facultad : Economía y Planificación	
Sumario	Sumarios (En, Es)	
Materia	<a href="#">INSTITUCIONES FINANCIERAS</a> <a href="#">CONSUMIDORES</a> <a href="#">COMPORTAMIENTO DEL CONSUMIDOR</a> <a href="#">METODOS ESTADISTICOS</a> <a href="#">MODELOS MATEMATICOS</a> <a href="#">ANALISIS DE LA REGRESION</a> <a href="#">TECNICAS DE PREDICCION</a> <a href="#">EVALUACION</a> <a href="#">PERU</a> <a href="#">ALGORITMO SMOTE</a> <a href="#">FUGA DE CLIENTES</a> <a href="#">RETENCION DE CLIENTES</a>	
Nº estándar	PE2018000144 B 7 M EUV E13; U10	

La retención de clientes ha tomado mucha importancia en los últimos años en las entidades financieras debido a la competencia agresiva por parte del sector, así como la autonomía del cliente en buscar mejores beneficios dentro de todas las ofertas que existen en el mercado bancario lo que se ve reflejado en el aumento de la tasa de clientes fugados. Ante esto se ha visto necesaria la implementación de técnicas estadísticas y/o técnicas de minería de datos, con la finalidad de

construir un clasificador predictivo que pueda ayudar a identificar a clientes potenciales a fugarse. En muchos casos cuando se aplican técnicas de clasificación, es común que la clase a predecir ocurra con menor frecuencia que la otra clase: la presencia de datos desbalanceados. Es decir, se tiene menor número de clientes fugados que no fugados, lo cual representa un inconveniente debido a que el clasificador necesita datos suficientes de ambas clases para poder aprender de ellas y así alcanzar una buena predicción. En esta investigación se propone el algoritmo Syntetic Minority Over-sampling Technique (SMOTE) como solución a este problema. SMOTE crea instancias nuevas a partir de un sobre-muestreo de las instancias existentes, llevando la clase minoritaria a un número suficiente para ser considerada balanceada y la clase mayoritaria si es necesaria reducirla mediante sub-muestreo aleatorio. En la presente investigación se validarán tales beneficios con la construcción de un modelo de regresión logística binaria con datos desbalanceados con y sin la aplicación del algoritmo de SMOTE; con el fin predecir la fuga de clientes en una entidad financiera. Se usarán para medir la precisión, la curva ROC y elementos de la comprobación de tabla cruzada como la especificidad y la sensibilidad.

## **Abstract**

Customer retention has taken much importance in recent years in financial institutions due to aggressive competition from the sector, as well as the autonomy of the client to seek better benefits within all offers that exist in the banking market, which is reflected in the increase in the rate of customers escaped. It has been necessary the implementation of statistical or technical techniques of data mining, in order to build a predictive classifier that can help identify potential customers to abscond. In many cases when classification techniques are applied, it is common to predict class to occur less frequently than other kind: the presence of unbalanced data. I.e. you have fewer customers escaped than not escapees, which represents a drawback since the classifier needs sufficient both kinds of data to be able to learn from them and thus achieve a good prediction. This research proposes the Syntetic Minority Over-sampling algorithm Technique (SMOTE) as a solution to this problem. SMOTE creates instances new starting from a sobre-muestreo of them instances existing, carrying the class minority to a number enough to be considered balanced and the class majority if is necessary reduce it by sub-sampling random. In the present study are validated such benefits with the construction of a model of binary logistic regression with unbalanced data with and without the application of the algorithm of SMOTE; in order to predict the flight of clients in a financial institution. They will be used to measure the precision, the ROC curve and elements of table cross as the specificity and sensitivity testing.