

## RESUMEN

**Autor** [Meza Rodríguez, A.](#)  
**Autor corporativo** [Universidad Nacional Agraria La Molina, Lima \(Peru\). Facultad de Economía y Planificación](#)  
**Título** **Estrategias comerciales para fortalecer la cadena productiva de la quinua orgánica en el distrito de Acobamba - Huancavelica**  
**Impreso** Lima : UNALM, 2018

### Copias

Ubicación	Código	Estado
Sala Tesis	<a href="#">E70. M49 - T</a>	USO EN SALA
<b>Descripción</b>	135 p. : 27 cuadros, 18 gráficos, 54 ref. Incluye CD ROM	
<b>Tesis</b>	Tesis (Ing Gestión Empresarial)	
<b>Bibliografía</b>	Facultad : Economía y Planificación	
<b>Sumario</b>	Sumarios (En, Es)	
<b>Materia</b>	<a href="#">AGRICULTURA ORGANICA</a> <a href="#">CHENOPODIUM QUINOA</a> <a href="#">TECNICAS DE MERCADEO</a> <a href="#">MERCADO</a> <a href="#">COMPETENCIA ECONOMICA</a> <a href="#">VENTAS</a> <a href="#">ZONA DE MONTAÑA</a> <a href="#">EVALUACION</a> <a href="#">PERU</a> <a href="#">QUINUA ORGANICA</a> <a href="#">ESTRATEGIAS COMERCIALES</a> <a href="#">CADENAS PRODUCTIVAS</a> <a href="#">ACOBAMBA (DIST)</a> <a href="#">ACOBAMBA (PROV)</a> <a href="#">HUANCAVELICA (DPTO)</a>	
Nº estándar	PE2018000734 B / M EUVZ E70	

La presente investigación tiene como propósito aplicar y comparar el modelo de regresión logística y el algoritmo Adaboost en datos desbalanceados, esto a efecto de predecir la fuga de clientes en una empresa del sector de telefonía móvil. El algoritmo Adaboost se sustenta en el aprendizaje adaptativo al entrenar clasificadores débiles combinándolos en conjunto para obtener un clasificador cuyo rendimiento sea fuerte. En cuanto a la regresión logística su modelamiento se realizó estrictamente desde una perspectiva de minería de datos, donde la clasificación es el objetivo y el rendimiento se evaluó en un conjunto de validación. Ambas técnicas se compararon mediante dos procedimientos, el primero mediante métodos de muestreo (sub-muestreo, sobre-muestreo y SMOTE) y el segundo modificando y/o ajustando el algoritmo o función. Al trabajar con datos desbalanceados la tasa de error de clasificación es ineficiente, por lo que las medidas de desempeño para elegir al mejor modelo fueron la precisión, el recall (sensibilidad), el F-measure, y como medida principal el AUC a través de curvas ROC. Al formar modelos logísticos con los métodos de muestreo, las medidas de desempeño arrojaron resultados similares, lo mismo pasó al formar modelos con el algoritmo Adaboost, sin embargo al comparar la regresión logística (AUC=0.86) con el algoritmo Adaboost (AUC =0.93), este último tuvo el mejor desempeño. En cuanto al ajuste a nivel de algoritmo o función, en la regresión logística se trabajó de dos maneras, el primero (Logit Asym) incluyendo en la FDA un valor Kappa ( $k$ ) y el segundo (Power Logit) un valor Lambda ( $\lambda$ ), en ambos modelos se identificaron los valores óptimos de  $k$  (0.02) y  $\lambda$  (2.5), en cuanto al algoritmo Adaboost (Adaboost Asym) se ajustó el

peso de la clase minoritaria cuyo costo de clasificación fue errónea. La comparación de estos tres modelos ajustados dio como mayor rendimiento al algoritmo Adaboost. Finalmente se realizó la validación cruzada con 10 iteraciones para todos los modelos dando resultados similares al método de retención. Realizada todas las comparaciones y las medidas de desempeño se concluye que el modelo óptimo para la predicción de fuga de clientes en la empresa de telefonía es el algoritmo Adaboost.

## **Abstract**

The purpose of this research is to apply and to compare the logistic regression model and the Adaboost algorithm in unbalanced data, the purposes of predict the customer churn in a company in the mobile telephony sector. The Adaboost algorithm is based on adaptive learning when training weak classifiers, combining them together to obtain a classifier whose performance is strong. In terms of logistic regression, its modeling was done strictly from a data mining perspective, where the classification is the objective and the performance was evaluated in a validation set. Both techniques were compared using two methods, the first using sampling methods (sub-sampling, oversampling and SMOTE) and the second modifying and / or adjusting the algorithm or function. When working with unbalanced data the classification error rate is inefficient, so the performance measures to choose the best model were accuracy, recall (sensitivity), F-measure, and as a main measure the AUC through ROC curves. When forming logistic models with the sampling methods, the performance measures yielded similar results, the same happened when forming models with the Adaboost algorithm, however when comparing the logistic regression (AUC = 0.86) with the Adaboost algorithm (AUC = 0.93), the latter had the best performance. Regarding the adjustment at the level of algorithm or function, the logistic regression was worked in two ways, the first (Logit Asym) including in the FDA a Kappa value ( $k$ ) and the second (Power Logit) a Lambda value ( $\lambda$ ), in both models the optimal values of  $k$  (0.02) and  $\lambda$  (2.5) were identified, in terms of the Adaboost algorithm (Adaboost Asym) the weight of the minority class whose cost of classification was erroneous was adjusted. The comparison of these three adjusted models gave the Adaboost algorithm a higher performance. Finally, cross validation was carried out with 10 iterations for all the models, giving similar results to the retention method. Once all the comparisons and measures of performance are concluded, it is concluded that the optimal model for the prediction of customer leakage in the telephone company is the Adaboost algorithm.