



Contribution ID: 22

Type: Abstract for Research Paper

Comparing Logistic Regression and Decision Tree Classifications Performance in the Context of Personal Cloud Storage Post-Adoption Behaviour

Machine learning literature is replete with algorithms for classification problems. The choice of an algorithm for a particular problem is not only dependent on statistical assumptions but also its performance. The current study compares the performance of logistic regression and decision trees when used in a binary classification in the context of personal cloud storage post-adoption behaviour. The users' intention to switch from freemium to premium personal cloud storage services was the classification problem. From literature review, six features were identified as predictors of intention to adopt premium personal cloud storage service. Data comprising the six features and a single dichotomous target was collected from university students. Machine learning techniques were used to balance the sample and split the data into training and validation sets. Classification analysis was then conducted on the data using both the logistic regression and decision tree algorithms. The performance of the classification algorithms was compared using the confusion matrix and the ROC Curve. For the decision tree, precision=0.75, recall=0.74 with an overall accuracy of 0.73 while for the logistic regression, precision=0.66, recall=0.65 with an overall accuracy of 0.65. The area under ROC curve for the decision tree was 0.79 while that of the logistic regression was 0.71. The decision tree algorithm therefore performed better than the logistic regression in all the metrics used for performance comparison. Perceived Usefulness, Perceived Risk and Perceived satisfaction emerged as the most important features in predicting users' propensity to migrate from freemium to premium personal cloud storage services.

Primary author: OREDO, John (University of Nairobi)

Track Classification: Data Science