

AWS Cost Management Adoption

Mark Russeff, Han Li, Andrea Hellebust, Hannah Khuong, Ying Xue, Keith Castelino
Seattle University – MGMT 5200

README

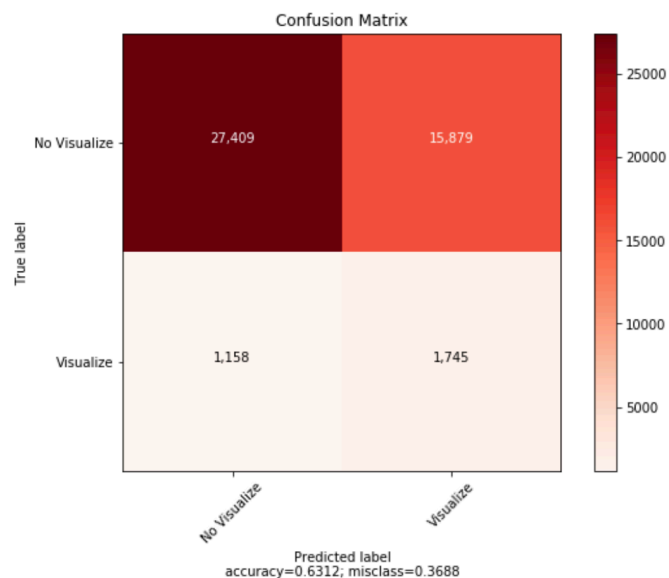
Our analysis is presented in a static EDA .html file, a static modelling .html file that walks through the logic of our programming and analysis and a .py file that can be run in order to examine output. For the .py file adoption/ billing files must all be in the working directory and the program can be run by an IDE of your choice using python3.

RESULTS

For our analysis we chose to use Random Forest Classification in order to predict whether a customer will use each of the AWS Cost Management Tools: Visualize, Report and Alert. While all of our models have good accuracy and recall numbers the precision of the Report and Alert models is extremely low. This is likely due to the imbalance in the data. Below you will find our results for our three models and our recommendations.

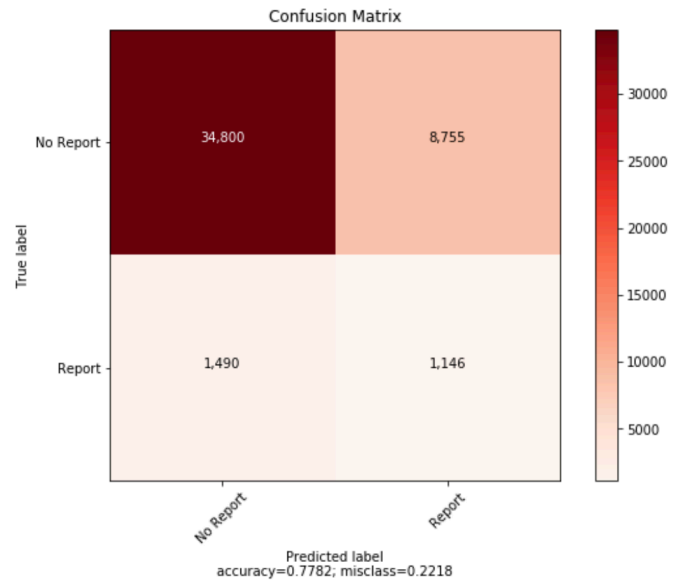
Visualize

Accuracy – 0.65
Recall – 0.56
Precision – 0.35
F1- score – 0.43



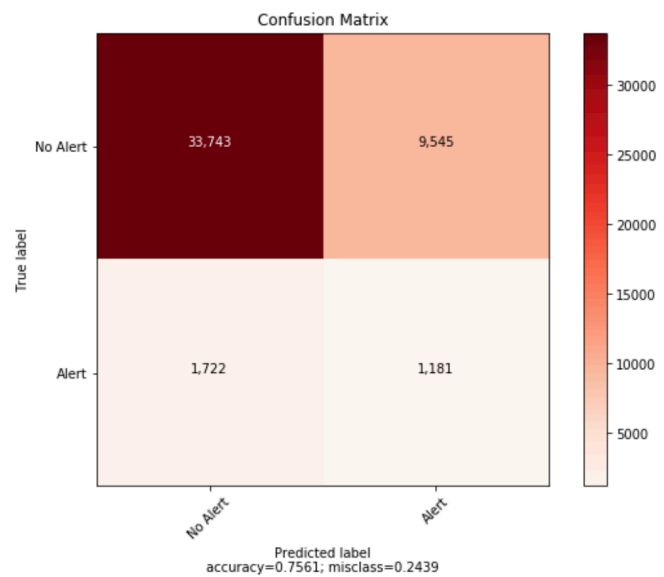
Alert

Accuracy – 0.78
Recall – 0.43
Precision – 0.12
F1- score – 0.18



Report

Accuracy – 0.76
Recall – 0.41
Precision – 0.11
F1- score – 0.17



The customer segmentation that was achieved by K-Means clustering showed that customers in cluster 4, those with a more diversified product mix, are more likely to use cost management and visualize in particular. This makes sound business sense as customers who use more products will have more challenges managing costs than those that use just one or two different products.

The models that we created do a fairly good job of predicting whether a customer will use an AWS Cost Management tool. However, due to the imbalance in the data, the models used to

predict Alert and Report are not very sound models. It would be easy to say that because they have high accuracy and recall that they are good models but because of the data imbalance they are also not very precise at all. With that being said the Visualize model does a very good job at predicting if a customer will use AWS Visualize or not. We would be comfortable using this model in production. Additionally, when we examine the output of the decision tree for visualize we can see that the number of products a customer used is the most important feature in our models. This makes sense because customers that use a lot of products will more than likely have challenges managing the cost of those products; challenges that AWS Cost Management can help with. Based on both of these model insights we can recommend that AWS focus thier marketing on customers who are likely to use visualize. These customers are likely from the market segment in cluster 4 and use more than 17 AWS products. By focusing on customers who use an even distribution of a lot of AWS products we are confident that AWS can grow the adoption rate of their Cost Management tools.