**CLASSIFICATION: PREDICTING CUSTOMER CHURN USING THE BOOSTED DECISION TREE IN AZURE MACHINE LEARNING**

**By Abusomwan Amadevborho**

Customer churn is usually one of the major problems in many businesses. Companies focus on building an accurate and reliable predictive model to identify customers that would churn soon. Azure ML is mostly used by businesses to explore and address complex analytical challenges of today and effortlessly scales to meet future needs. As seen in Figure 1.0 the CRM dataset shared went through various components or operators in the pipelines to get the best model for predicting customer churn. A noticeable interesting feature about Azure is timesaving when making predictions in the cloud, hence do not have to rerun every operator or component unlike other cloud-enabled machine learning tool for example SAS Viya. This feature is a positive aspect of working with Azure ML. Nevertheless, running a pipeline on the cloud can also be time-wasting as the component can also take a longer time to run since the deployment is happening in a data centre and everyone else is also using it. The smote component as seen in Fig 1.0 was useful to increase the number of underrepresented cases in a dataset that's used for machine learning. Hence, giving 70% in training the model with the Two-Class Boosted Decision Tree before scoring and evaluating the model. This is another good feature because we can see “Train Model” has a visible operator which is easy to understand unlike the usual case where training and testing is within the “Split Data”. The outcome using the Two-Class Boosted Decision Tree can be in seen Fig 1.1 whereby we have a c-shaped ROC and a good AUC and accuracy.

Furthermore, the prediction of flight delay pipeline as seen in Fig 2 done by my team member. The entire runtime which took about 16minutes is another aspect of Azure ML in which resource group and workspace deployment is done on the cloud, hence usually takes a long time for a prediction. Also, another interesting feature was the pipeline flow where the weather dataset and flight delay data passed different components in the pipeline like cleaning and execution of the R script before the data was joined by the “Join Data” operator to get a model prediction using a Two-Class Logistic regression model. Lastly, looking at a different case study for Regression model: Comparing algorithms for Price Prediction as seen in Fig 3.0 Where we have a model comparison using both the decision forest regression and Boosted decision tree regression to get the champion model where results show that the Boosted Decision Tree Regression model has a lower root mean squared error than Decision Forest Regression model.

**BUILDING A CLASSIFIER TO PREDICT FLIGHT DELAYS WITH AZURE MACHINE LEARNING DESIGNER WITH INPUT OF R SCRIPT**

**BY AJULO JOSEPH**

Azure machine learning studio was used to build a classifier which R script was used to predict flight delays using a flight historical and weather dataset. The target label is to predict classes: delayed, or on time. Azure ML allows the user to pre-process each dataset separately. Also, the Azure ML pipeline supported the use of R script execution on the flight and weather data to round the departure time and observation time. This feature shows how powerful and flexible Azure ML pipeline. Hence, after both datasets are joined the Two-Class Logistic Regression model was used to train the data. The entire runtime of the pipeline took about 16mins which was fair enough. The quality of the model was accessed by the execute model module. The output charts and tables of the classification model to check its accuracy and performance were limited. Therefore, Azure machine learning provided limited information about the built classifiers unlike other point and click machines tools like RapidMiner or SAS. The pipeline and result of the experiment using Azure ML is available in fig (2).

A customer churn business dataset where the Two-Class Boosted Decision Tree was used to train the data done by team member as seen in Fig 1.0. From his experiment Azure ML proves to be very effective and fast in running the pipeline effectively, hence another interesting feature of the customer churn pipeline is the Smote module which basically increase the number of underrepresented cases of the data before training the model. Therefore, the model with smote operator is more effective. Finally, looking at the regression model where my team member compared algorithms for Price Prediction as seen in Fig 3.0 A model comparison was made using two separate models where the coefficient of determination result was given as 0.918 shows it’s a better model.

**Regression: Comparing algorithms for Price Prediction**

**BY IDOWU OMONIYI JAMES**

The Azure machine learning studio is a particularly good programme for regression building and, in this case, there is going to be a comparison for of regression model to correctly predict prices. From the Figure below the data is after being cleaned is split into 2 models. These are the models which are being compared in this instance. We have the Boosted Decision Tree Regression and the Decision Forest Regression; both use slightly different methods but still come to a similar result set. It is shown from the fig 3.0 that after the data is split a Train Model and Score model is created to compare both results. The results can be seen in the Fig 3.1 below showing the comparison between Mean Absolute Error between the left position and right position. The Coefficient of Determination from the figure below was also compared between the two models, with both providing a slightly greater than 0.9 value, thus practically indicating an above 90% reliability of the dependable variable meaning it is exceptionally reliable and can be used for further predictions.

Azure studio has proven to be an effective mode for various operations ranging from Predictions to Classifications, but it has not been without its flaws; for instance, from this graph pipeline and the previous explained by my team members, the processing time has always been some sort of a disappointing factor. Since Azure remains a cloud-based program, submitting and running various can be time demanding and any form of improvement on these services will always come at a higher cost making it costly and high maintenance. In terms of the technical aspects, there is a slight limitation to access of the pipeline being ran. For example, after running and successfully completing the pipeline an Analyst will unfortunately not be able examine the ‘Coefficient Weights’ or ‘Y-Intercept’ of a Model very easily.

**Appendix 1**

**Graphical user interface, application

Description automatically generated**

*Fig 1.0* Customer Churn Azure ML pipeline

**A picture containing text

Description automatically generatedChart, line chart

Description automatically generated**

*Fig 1.1 ROC* curve and Outcome

**Appendix 2**

**Diagram

Description automatically generated**

Fig (2a) Pipeline showing modules used

**A picture containing graphical user interface

Description automatically generated**

**Chart, line chart

Description automatically generated**

Fig(2b) Results from the Evaluation of the model.

**Appendix 3**

**Diagram

Description automatically generated**

Fig 3.0: *Showing the Graph Pipeline with two Separate Regression Models*

**Graphical user interface, text, application, email

Description automatically generated**

Fig 3.1: *Showing The Metric Values and Comparing both Models*