# Entregable obligatorio Modulo Machine Learning

# (Grupo 340 / Subgrupo 6)

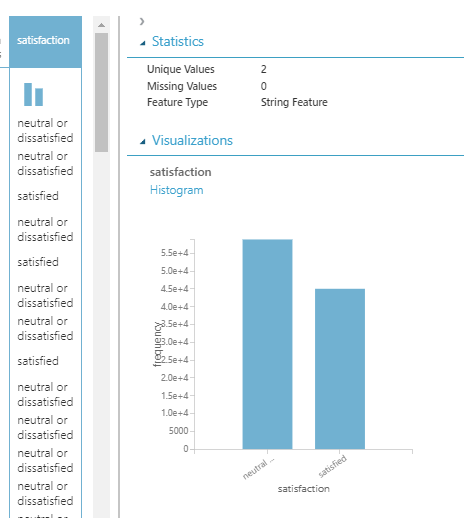
## Nombre de Integrantes y link al experimento

Nombre de los integrantes: Catalina Banchieri, Nicolás Olivera, Martín Sacco García.

Link al experimento: <https://gallery.cortanaintelligence.com/Experiment/Obligatorio-Machine-Learning-Grupo-6-Copy>

## Definición del problema:

Dada la variable que se quiere predecir, que estaría siendo la satisfacción del cliente; la misma según los datos del Dataset, es del tipo binaria, con valores de salida “satisfied” O “Neutral or dissatisfied”. Se considera binaria dada que la variable objetivo tiene 2 valores posibles de salida. También cabe destacar que estaría siendo un tipo de problema supervisado.



Debido al detalle mencionado, se considera entonces que estamos ante un problema del tipo “Variable Dependiente Binaria” a trabajar a través de este modelo.

## Determinar qué métrica se va usar y porque:

La métrica principal a utilizar será el “Accurancy” la cual demuestra la eficacia del modelo para la predicción buscada. También se analizarán los valores de “True Positive”, “True Negative”, “False Positive” y “False Negative”, así como también la comparativa relativa entre las gráficas y el área bajo la curva (AUC). El modelo seleccionado será el de mayor “Accurancy”.

## Análisis de variables encontradas y a su vez ideas de transformaciones necesarias u opcionales

Para el análisis de variables, se adjunta Excel con el mismo en la pestaña “Variables”.



**Observación:** El análisis de variables encontradas se realiza teniendo en cuenta el significado de la variable establecida para el dataset. (Ejemplo, si tengo pocos valores únicos respecto a la cantidad de filas del Dataset, estoy ante una variable discreta).

**Observación:** Dentro del análisis de variables se detectó la variable “Arrival delay in Minutes” con datos faltantes y se completan los faltantes con la mediana.

## Selección del modelo a usar

Se seleccionó el modelo Two-Class Boosted Decision Tree debido a su capacidad para modelar con datos binarios.

## Interpretación de los coeficientes del modelo usado (del que de mejor)

Para el análisis de modelos, se adjunta Excel con el mismo en la pestaña “Comparativa entre modelos”.



## Conclusión

El primer modelo es el mejor. Si bien las agrupaciones o la ingeniería de variables es una buena práctica y recomendad para ciertos casos, bastó con las transformaciones iniciales como para, con el algoritmo seleccionado y las métricas definidas, obtener un resultado considerado aceptable desde el punto de vista absoluto sobre el mismos; así como también relativo entre los modelos y sus diferencias.