Machine Learning Canvas

Proyecto: Windows Malware Prediction **Autor:** Alfredo Mariño **Fecha:** 19/02/2022 **Version:** 1.0

<table>
<tr><td>NUEVOS DATOS Y REENTRENAMIENTO

Mensualmente se recopilaran los datos de deteccion de virus en los equipos windows para reentrenar el modelo</td><td>PREDICCIÓN (ON / OFF)

se haran las predicion en batch una vez al mes</td><td rowspan="2">PROPUESTA DE VALOR

Se requiere predecir cuando un computador que usa el sistema operativo Windows es vulnerable o propenso a ser infectado por un software malicioso (Malware).</td><td>ORÍGENES DE DATOS

Muestra de 500.000 registros donde cada fila del dataset corresponde a una máquina única. Estos datos provienen del dataset de la competición de Kaggle Microsoft Malware Prediction y se basan en las características obtenidas en la solución de endpoint Windows Defender</td><td>TAREA DE ML

Clasificacion supervisada</td></tr>
<tr><td rowspan="2">EVALUACIÓN EN SERVICIO Y ALM

Se iran contrastando las predicciones del modelo contra computadores que tengan las mismas caracteristicas con las que se creo el arbol de decision</td><td>MÉTRICA DE EVALUACIÓN (EN DESARROLLO)

Se utliza el AUC y los accuracy</td><td>ATRIBUTOS

Algunas de las variables mas relevantes son: AVProductsInstalled, AVProductsEnabled, IsProtected, Wdft_IsGamer, SmartScreen</td><td rowspan="2">DEFINICIÓN DEL PERÍMETRO Y TARGET (SÓLO EN CS)

El target es la variable booleana HasDetections que establece si ese equipo fue infectado por un malware, si=1 o no=0</td></tr>
<tr><td colspan="3">USO DEL MODELO, TOMA DE DECISIONES Y EXPLICABILIDAD

El modelo se usará para implementarlo en la aplicación Windows Defender dando aviso al usuario cuando su máquina supere un cierto umbral de probabilidad de ser infectada.</td></tr>
</table>