***Trabajo Sesgos en IA***

***Preguntas***

***Tarea #1 de FairAware***

¿Hay valores de características faltantes para un gran número de observaciones?

* No se presentan datos nulos para ninguna de las columnas del dataset

¿Hay funciones que faltan que podrían afectar otras funciones?

* No hay funciones faltantes dentro del dataset trabajado.

¿Hay algún valor de característica inesperado?

* Al analizar el comportamiento de los datos del dataset se evidencia que se presentan algunos valores alejados del promedio en algunas columnas del dataset, esto ocurre por ejemplo en la variable de distancia de la compra desde casa (distance\_from\_home) o en la variable de distancia desde la última compra (distance\_from\_last\_transaction), al final estas variables pueden contribuir a mejorar la precisión del modelo al realizar predicciones cuando se presentan este tipo de comportamientos.

¿Qué signos de sesgo de datos ve?

* Se puede observar sesgo en las siguientes variables de entrada:

**repeat\_retailer:**  Esta variable de entrada tiene dos posibles valores, 0 cuando la transacción realizada se hace en una tienda donde no se se había comprado antes y 1 cuando la compra se hace en una tienda en la que anteriormente ya se había comprado, al observar la distribución de los datos indica que tan solo el 11.84% de las transacciones se realizan en tiendas en las que anteriormente no se había comprado, por lo que es posible incurrir en un sesgo de reporte, en el que la distribución de los valores en el conjunto de datos no representan con exactitud la distribución de estos datos en el mundo real.

**used\_pin\_number:** Esta variable de entrada puede tener dos posibles valores, 0 para indicar que la transacción no hizo uso del pin numérico y 1 para indicar que la transacción se realizó con pin numérico, al observar la distribución de los registros en este variable se evidencia que cerca del 90% de las transacciones no hacen uso del pin numérico, lo que puede entrar en contradicción con el porcentaje de compras online cercano al 35%, pues este tipo de capas de seguridad adicionales se usa generalmente al realizar transacciones de forma física, es posible que se esté incurriendo en un sesgo de reporte al no representar con exactitud la distribución de los datos en el mundo real.

***Tarea #2 de FairAware***

¿Qué falta?

¿Qué se está sobregeneralizado?

* Se sobregeneraliza el fraude cuando no es una compra en una tienda repetida y se usa el pin number.

¿Qué es estar subrepresentado?

¿Cómo reflejan las variables y sus valores el mundo real?

al utilizar la herramienta creada para ver la distribución de los datos nos dimos cuenta que al plotear la compra en una misma tienda vs uso de pin numérico evaluando el fraude que es la salida tenemos que el modelo nunca detecta que se genera fraude cuando se compra en un lugar nuevo con el uso de pin numérico, que en la vida real tiende a ser lo contrario.

¿Qué podríamos estar dejando de lado?

Generar una mayor cantidad de datos para procesos de compras con tarjeta de crédito por distancia de transacción y distancia de la casa, debido a que la distribución de datos actual refleja datos atípicos que pueden generar sesgos en el modelo.

***Tarea #3 de FairAware***

¿Cuál de las siguientes características podría plantear un problema con respecto al online orden?

Elija una característica de las opciones desplegables en la celda a continuación y luego ejecute la celda para verificar su respuesta. Luego explore el resto de las opciones para obtener más información sobre cómo cada una influye en las predicciones del modelo.

***Tarea #4 de FairAware***

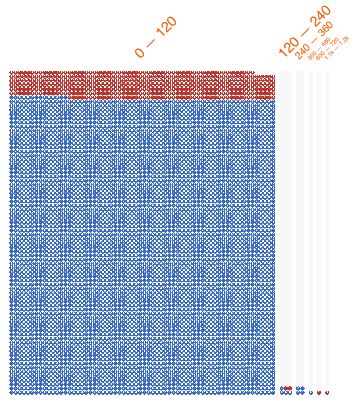
¿Existen disparidades significativas en las tasas de error que sugieran que el modelo funciona mejor para un subgrupo que para otro?

————————————————————————————————————————

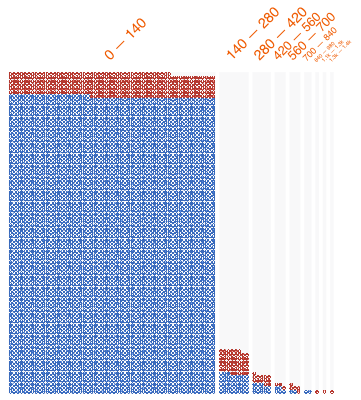
**PARA LA PRESENTACIÓN**

**Datos atípicos:**

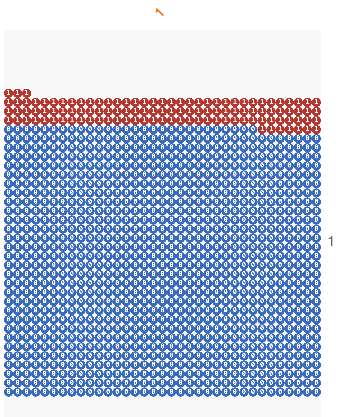
Distancia desde la ultima compra:



Distancia desde casa:



Compras online con uso de chip(física):



**Datos que no representan en realidad la distribución del mundo real: (sesgo de reporte)**

Compras en tienda nueva y haciendo uso de Chip

