

# VARIABLES DE ASEGURADORA DE SINIESTROS

VARIABLE	DESCRIPCIÓN	RANGO DE VALORES
Morosidad1	Tasa de morosidad para Siniestro1	"Sin deuda"
Morosidad2	Tasa de morosidad para Siniestro2	"1-29 días"
Morosidad3	Tasa de morosidad para Siniestro3	"30-59 días"
Morosidad4	Tasa de morosidad para Siniestro4	"60-89 días"
Morosidad5	Tasa de morosidad para Siniestro5	"90-119 días"
Morosidad6	Tasa de morosidad para Siniestro 6	"120-149días", "150-179 días"
Siniestro1	Accidente de auto	
Siniestro2	Daños en propiedad	
Siniestro3	Enfermedades cubiertas	
Siniestro4	Robos	
Siniestro5	Incendios	
Siniestro6	Daños por clima	
Estado Siniestro	Estado del siniestro (victima o no)	0 = No resuelto 1 = Resuelto
Nivel Ingresos	Nivel de ingresos del cliente	0, 1, 2, 3, 4, 5
Antigüedad_Maxima	Antigüedad del cliente	Número entero
Saldo pendiente tota	Saldo del cliente que debe a la aseguradora	Número irracional
Saldo pendiente de la aseguradora	Saldo que debe la aseguradora al cliente que sufrieron algún tipo de siniestro.	Número irracional

## PROCEDIMIENTO DEL TRATAMIENTO.

- ✓ Se transformo las variables a su rigén según el diccionario de variables
- ✓ Se imputo las variables cualitativas por la moda
- ✓ Se imputo los na de variable antigüedad por la mediana dada su distribución asimétrica y sesgo a la derecha.
- ✓ Se trato los outliers suavizo, teniendo en cuenta que si es posibles valores altos den la data de una aseguradora. El método que se utilizo es winscored, no se suavizó con logaritmo, ni raíces y otro tipo que también podría ser adecuado cuando hay valores elevados. No se usé la moda, ni la media dado que son valores cero y anularía la existencia de dichas variables al ser todo cero en los diferentes tipos de saldos, lo cual no tendría ninguna lógica.
- ✓ Se limpio la data de saldo pendiente que no es del seguro en función a la tasa de morosidad si no hay morosidad el cliente no debe tener saldo pendiente.
- ✓ No se escogió método robusto para tratar os outliers dado que al predecir valores tenía inconsistencia, por ejemplo, si predecía valores de saldos pendiente de la aseguradora en función a estado siniestro 1,2,3,4,5,6 cambiaba los valores de manera ilógica le ponía un saldo diferente cero a un individuo que nunca sufrió siniestro, si no tubo siniestro su saldo debería de ser cero. Recordar que bajo esta última condicional también se limpió la data.
- ✓ No se usó árbol de decisión para tratar los outliers dado que no los modifican si no que los mueven a distintos lugares a diferentes observaciones según la data y las otras variables, lo cual podría modificar de manera errónea la información de un ay otras observaciones, dado que consideramos que la data los tiene algunos errores de

digitación y no de organización del llenado de datos en sus filas y columnas, no lo segmentamos con el árbol.

- ✓ Siniestro no se trato con median y media por ser cero o cercano a ello, ya que si fuera así desapareceríamos a la variable con puros ceros, e trato con winscorización para quitar ese extremo valor de 289.
- ✓ Se procedió a realizar la prueba de normalidad y a no ser normales, se procedió a realizar pruebas no paramétricas.
- ✓ Se necesita realizar un modelo logit y probit para determinar la probabilidad de siniestro en función a nuestras variables.
- ✓ Se debe agregar más variables como edad, sexo, industria laboral, estado de salud, etc., según un arco teórico y conceptual que tenga relación con siniestros.