



11-1-2025

# TAREA: CREDIT SCORING\_TELCO

GESTIÓN GLOBAL DE RIESGO-SCORING

## GRUPO A

**XINYUAN ZHENG**

xinyuanz@ucm.es



## ÍNDICE

1.	INTRODUCCIÓN .....	2
1.1.	Contexto del Problema .....	2
1.2.	Objetivo del Proyecto.....	2
2.	ANÁLISIS EXPLORATORIO .....	2
2.1.	Conjuntos de Datos Utilizados.....	2
2.2.	Resumen de Variables Clave .....	3
2.3.	Preparación de los Datos .....	3
2.3.1.	Limpieza de Datos .....	4
2.3.2.	Transformaciones y Selección de Variables .....	5
3.	MODELOS .....	7
3.1.	Árbol de decisión .....	7
3.2.	Modelo de regresión .....	8
4.	TARJETA DE PUNTUACIÓN .....	11
5.	RESULTADOS DEL TEST CIEGO .....	14



---

## 1. INTRODUCCIÓN

---

### 1.1. Contexto del Problema

---

MovilTV es una empresa que ha experimentado un crecimiento acelerado en el mercado español durante los últimos cinco años. Sin embargo, la creciente cartera de clientes también ha incrementado el riesgo asociado a impagos, los cuales generan un impacto negativo significativo en la rentabilidad. La compañía estima que cada impago equivale al valor económico de 20 nuevas contrataciones. Este desafío ha llevado a la empresa a buscar una solución que permita identificar y priorizar clientes con alta probabilidad de cumplimiento de pago.

### 1.2. Objetivo del Proyecto

---

El principal objetivo de este proyecto es desarrollar una estrategia predictiva que permita identificar clientes con alta probabilidad de incumplir con sus pagos, mediante la construcción de un modelo de scoring traducido a una tarjeta de puntuación. Este modelo será evaluado a través de un test ciego, asegurando que solo se otorguen créditos a clientes con bajo riesgo de impago, optimizando así el proceso de captación y reduciendo pérdidas asociadas.

---

## 2. ANÁLISIS EXPLORATORIO

---

### 2.1. Conjuntos de Datos Utilizados

---

Para realizar este trabajo, se ha tenido acceso a tres conjuntos de datos principales:

- TrainCS: Clientes con Comportamiento Conocido

Este dataset contiene 100,061 observaciones y 73 variables, que incluyen información demográfica, financiera y el comportamiento de pago (clasificado en pagadores e impagadores). Este conjunto será utilizado para el entrenamiento y validación del modelo, dividiendo los datos en un 70% para entrenamiento y un 30% para validación, lo que permitirá ajustar y evaluar el modelo de manera adecuada.

- CS\_Rejected: Clientes Rechazados

Este dataset incluye 29,109 observaciones y 74 variables; recopila información de clientes rechazados por las políticas de riesgo históricas de la compañía. Aunque no contiene etiquetas de comportamiento de pago, su uso se centrará en el análisis exploratorio para comparar características de clientes rechazados y aceptados, así como en la verificación de la coherencia entre el modelo propuesto y las decisiones histórica.

- Scoring\_CS: Dataset Ciego

Este conjunto de datos consta de 66,647 observaciones y 72 variables, y está diseñado para probar el modelo final. Al no incluir información sobre el comportamiento de pago, se utilizará en el último paso del análisis para aplicar el modelo entrenado, generar las puntuaciones y clasificar a los clientes en dos categorías: “Aprobados (0)”, correspondientes a aquellos con alta probabilidad de pagar, y “Rechazados (1)”, para clientes con alto riesgo de impago.

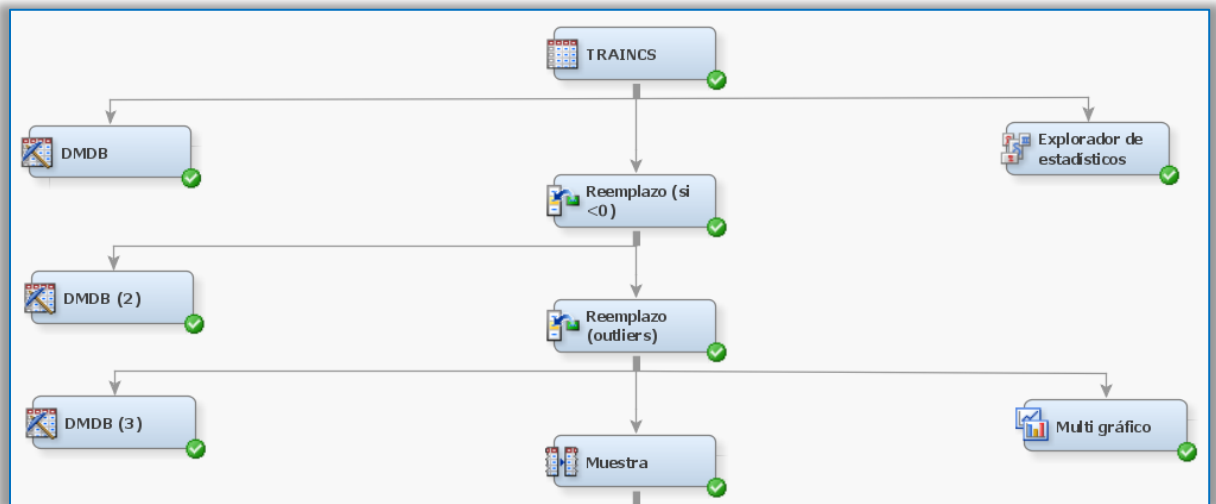
## 2.2. Resumen de Variables Clave

Los conjuntos de datos proporcionados contienen una amplia variedad de variables relacionadas con aspectos demográficos, financieros, de comportamiento y uso de servicios. Estas variables son clave para analizar los factores asociados al abandono temprano de clientes, representado por la variable dependiente “Target\_i\_ECInvol\_12m”.

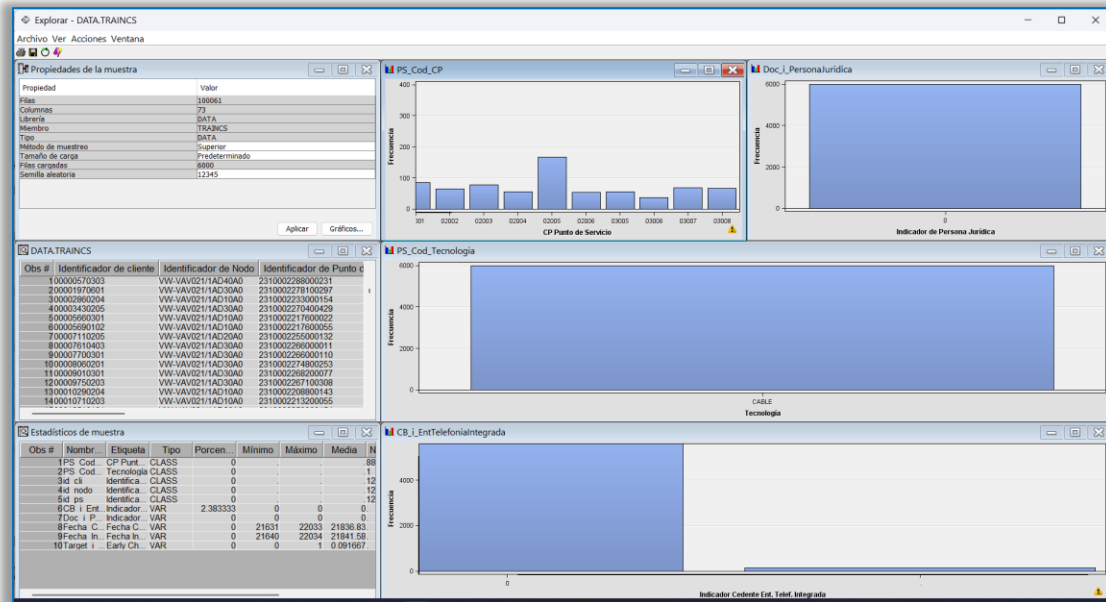
Esta variable indica si un cliente ha abandonado en los últimos 12 meses, tomando el valor “1” si el cliente ha abandonado (alto riesgo de churn) o “0” si el cliente no ha abandonado (bajo riesgo de churn). Por lo tanto, será utilizada como la variable objetivo (target) para el modelo predictivo. El resto de las variables se incluirá en el conjunto de datos predictoras “input”.

## 2.3. Preparación de los Datos

El flujo de los procesos realizados para la preparación previo a la construcción del modelo de tarjeta de puntuación para el análisis, se muestra en la siguiente imagen.



Antes de empezar con las fases de preparación de los datos, tras importar el dataset “TRAINCS” se ha realizado un ajuste en el tipo de variable según los datos que muestra cada variable, además de una exploración previa de cada tipo de categoría de variables para detectar aquellas variables inadecuadas para el modelo.



En principio se ha descartado la variable “PS\_Cod\_CP” debido a que sobrepasa el límite de categorías permitida para ejecutar el nodo, y al tratarse de una variable que determina el código postal, existe otra variable que representa la provincia, por lo tanto se ha decidido descartar esta variable. Por otra parte, las variables “Doc\_i\_PersonaJuridica”, “PS\_Cod\_Tecnologia” “CB\_I\_EntTelefonialIntegrada” al ser variables binarias o nominal y contar con menos de 2 categorías, se considera variables incompletos y poco representativos para el análisis, y es por ello por lo que también se descartan.

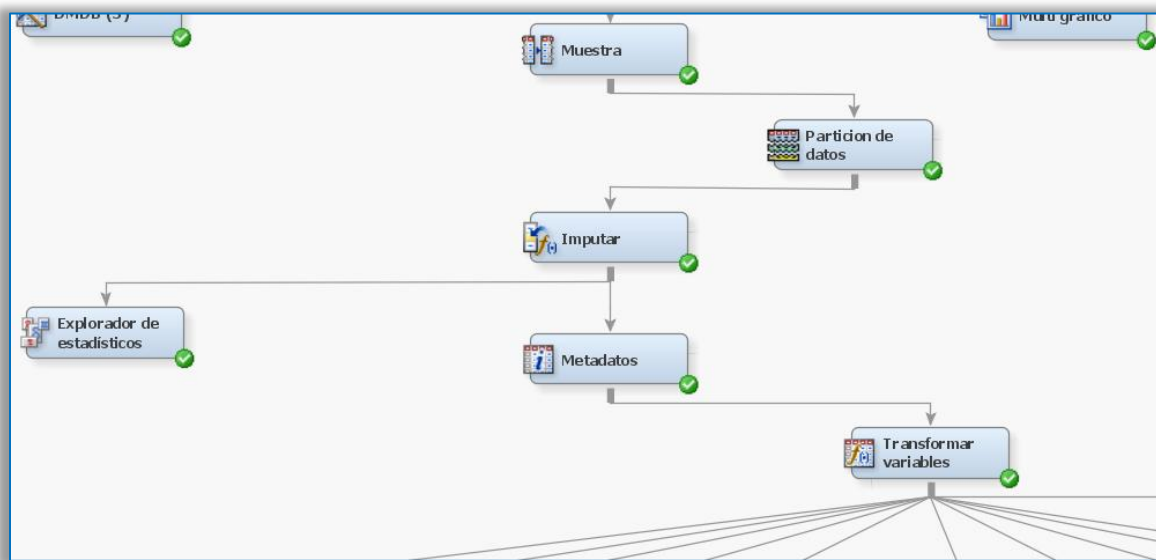
### 2.3.1. Limpieza de Datos

Variable	Etiqueta	Ausente	N	Mínimo	Máximo	Media	Desviación estándar	Asimetría	Curtosis
CR_Int_DeudaDescubiertoCC	Deuda por descubierto en C/C	2827	97234	0.00	9277854.00	1032.95	38656.87	165.133	35876.53
CR_Int_DeudaMedia	Deuda Media (Bureau Credito)	2827	97234	0.00	155655035.00	35386.65	709330.83	131.684	25605.72
CR_Int_DeudaMoTelco	Deuda Mo Telco	2827	97234	0.00	253617857.00	89248.46	1653171.91	83.893	10249.17
CR_Int_DeudaPrestatipo	Deuda por un préstamo hipotecario	2827	97234	0.00	205440626.00	39675.61	1388964.28	96.737	12047.13
CR_Int_DeudaPrestPerso	Deuda por un préstamo personal	2827	97234	0.00	70630097.00	17047.16	466869.55	81.087	9074.51
CR_Int_DeudaPresttao	Deuda por un préstamo	2827	97234	0.00	252911127.00	56722.77	1570456.15	95.576	12453.76
CR_Int_DeudaTarjetas	Deuda por tarjetas	2827	97234	0.00	11197778.00	7941.73	96353.91	37.506	2732.23
CR_Int_DeudaTelco	Deuda Telco	2827	97234	0.00	1277960.00	3801.54	20139.97	14.514	448.73
CR_Int_DeudaTotal	Deuda Total	2827	97234	0.00	253631457.00	93049.99	1654428.50	83.740	10221.02
CR_Int_MaxSaldoImpagado	Max. Saldo Impagado	2827	97234	0.00	199209451.00	68630.30	1180704.12	86.362	11940.96
CR_Nua_Cedentes	Número Cedentes	2827	97234	0.00	42.00	0.25	0.92	7.138	104.31
CR_Nua_DiasCB	Días desde que entró en Bureau Credito hasta la consulta	2827	97234	0.00	2040.00	97.61	319.05	3.504	11.60
CR_Nua_DiasBdUltiasDeuda	Días desde última deuda	2827	97234	0.00	1963.00	71.05	254.38	4.138	17.43
CR_Nua_DiasEntreDeudas	Días entre deudas	2827	97234	0.00	1885.00	26.56	149.12	6.935	52.63
CR_Tasa_TelcoTotal	Tasa Telco-Total	2827	97234	0.00	1.00	0.07	0.24	3.494	10.35
Dda_Int_DeudaCli	Deuda Cliente	0	100061	-243.32	28743.52	0.40	91.33	311.656	98055.28
Dda_Int_DeudaPS	Deuda Punto de Servicio	0	100061	-1531.72	14042.82	18.31	129.51	32.782	2492.41
PS_Cod_SecCensal	Código Censal	942	99119	0.00	5029711019.00	2503133508.60	1577169987.68	-0.149	-1.22
SD_Int_GtoAlqTV	Gastos Alquiler TV	26136	73925	1.03	531.23	70.15	36.36	1.794	9.16
SD_Int_GtoArar	Gastos Arar	26136	73925	0.00	133.40	2.22	5.08	9.964	168.55
SD_Int_GtoCineTeatro	Gastos Cine-Teatro	26136	73925	0.00	192.22	8.35	10.75	4.934	45.48
SD_Int_GtoCineTelef	Gastos Equipos Teléfono	26136	73925	1.16	189.25	12.50	6.52	4.366	76.66
SD_Int_GtoFiestas	Gastos Fiestas	26136	73925	11.36	666.60	99.56	46.21	1.463	4.19
SD_Int_GtoLibrosText	Gastos Libros Texto	26136	73925	41.98	2783.06	415.35	203.35	1.133	2.63
SD_Int_GtoMaterialInf	Gastos Material Informacion	26136	73925	5.22	501.42	91.37	34.75	1.433	5.35
SD_Int_GtoRestoLibro	Gastos Resto Libros	26136	73925	1.95	377.16	38.80	22.88	2.433	12.91
SD_Int_GtoTV	Gastos TV	26136	73925	16.73	414.31	106.66	32.57	1.177	4.11
SD_Int_GtoTelefMovil	Gastos Teléfono Móvil	26136	73925	199.17	1055.02	486.64	93.34	0.928	1.84
SD_Int_GtoVacacionT	Gastos Vacaciones Todo Incluido	26136	73925	1.62	538.29	79.23	48.82	1.861	6.38
SD_Int_IngrActivos	Media Ingresos Netos Familias Activas	26136	73925	1288.91	4833.34	2469.49	411.28	0.798	0.91
SD_Int_IngrNetos	Media Ingresos Netos Familias	26136	73925	1039.75	4061.93	2214.20	391.61	0.682	0.63
SD_Int_IngrPensionistas	Media Ingresos Netos Fam. Pensionistas	26136	73925	690.71	5289.70	1967.40	434.54	0.884	1.81
SD_Int_IngrXPers	Media Ingresos por Persona Activa	26136	73925	582.28	2528.82	1378.92	211.79	0.347	0.41
SD_Tasa_HabExtr	Tasa Habitantes Extranjeros	26136	73925	0.00	0.72	0.12	0.09	1.228	1.70

Para el proceso de limpieza de los datos, se utiliza el nodo de “DBDM” para encontrar aquellas variables que presenta valores negativos o atípicos. Así como las variables con valores ausentes (NA) que se tratara posteriormente con la imputación.

Una vez determinado la presencia de estos datos, se procede a usar el nodo de “Reemplazo” para transformar los valores negativos y los outliers en valores NAs que será tratado posteriormente.

### 2.3.2. Transformaciones y Selección de Variables



Una vez limpiado los NAs y los outliers, se procede con estas fases previo a la construcción del modelo.

Datos=DATA					
Variable	Valor numérico	Valor formateado	Número de ocurrencias	Porcentaje	Etiqueta
Target_i_ECInvol_12m	0	0	89750	89.6953	Early Churn Involuntario 12 meses
Target_i_ECInvol_12m	1	1	10311	10.3047	Early Churn Involuntario 12 meses
Datos=SAMPLE					
Variable	Valor numérico	Valor formateado	Número de ocurrencias	Porcentaje	Etiqueta
Target_i_ECInvol_12m	0	0	10311	50	Early Churn Involuntario 12 meses
Target_i_ECInvol_12m	1	1	10311	50	Early Churn Involuntario 12 meses

Al estudiar la distribución de las categorías de la variable objetivo “Target\_i\_ECInvol\_12m”, la proporción de clientes abandonado (1) es del 10,30% mientras que la proporción de no abandonado (0) es de 89,69%.

Como el objetivo es analizar los clientes con alto riesgo de impago (1), esta descompensación entre las 2 categorías, podría inducir a generar un modelo desbalanceado. Por lo tanto, se procede a igualar el peso entre las 2 categorías mediante un sobremuestreo para que sean 50-50.

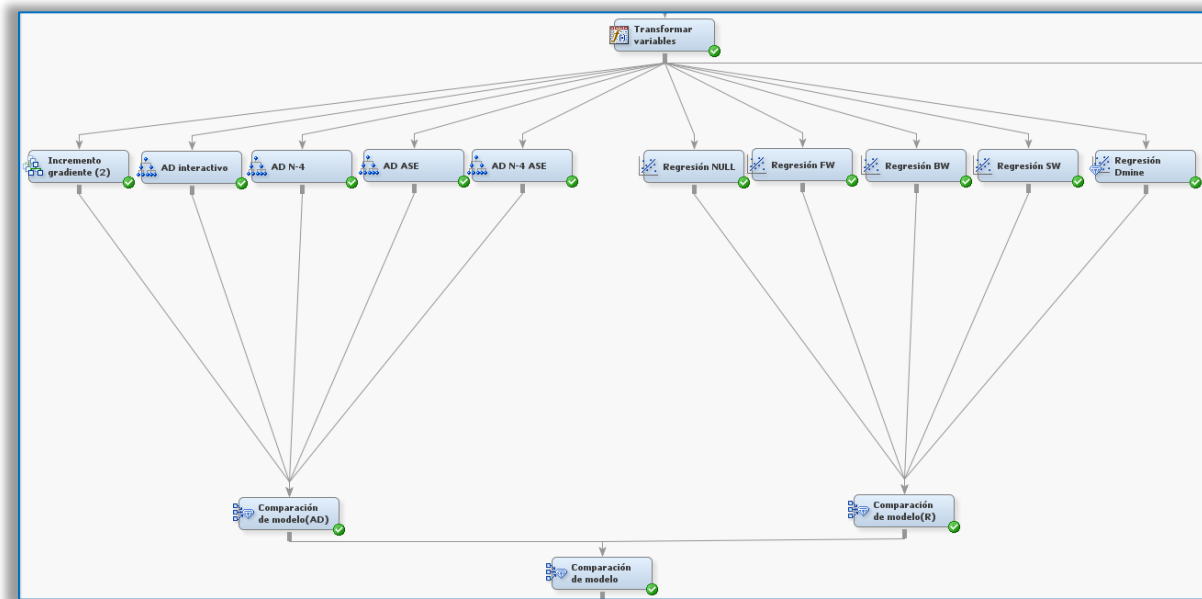
Luego, a la hora de particionar la muestra, debido a que es un “dataset” grande, se ha establecido el criterio de 70% para entrenamiento y 30% para la validación.

.. Propiedad	Valor
<b>General</b>	
ID de nodo	Impt3
Datos importados	...
Datos exportados	...
Notas	...
<b>Entrenamiento</b>	
Variables	...
Variables no ausentes	No
Corte ausente	50.0
<b>Variables de clase</b>	
Método de imputación p	Distribución
Método predeterminado	Ninguno
Normalizar valores	Sí
<b>Variables de intervalo</b>	
Método de imputación p	Mediana
Método predeterminado	Ninguno

Debido a la gran cantidad de valores ausentes, se procede a utilizar el nodo de “Imputar” para sustituir estos valores. Se ha establecido el criterio de corte en las variables con un max de 50% de NA. Para las variables tipo clase (Nominal) se ha usado el método de “distribución” mediante un proceso aleatorio de asignar esos valores faltantes con una categoría presente en la variable. Y para las variables, de intervalo, se ha usado la mediana para sustituir esos NA.

Una vez imputados los datos, se crea el “Metadatos” para dejar almacenado las características de las variables antes de la construcción de los modelos, y se crea el nodo “Transformar variables” para revisar aquellas variables que se requiera alguna modificación para mejorar el resultado.

### 3. MODELOS



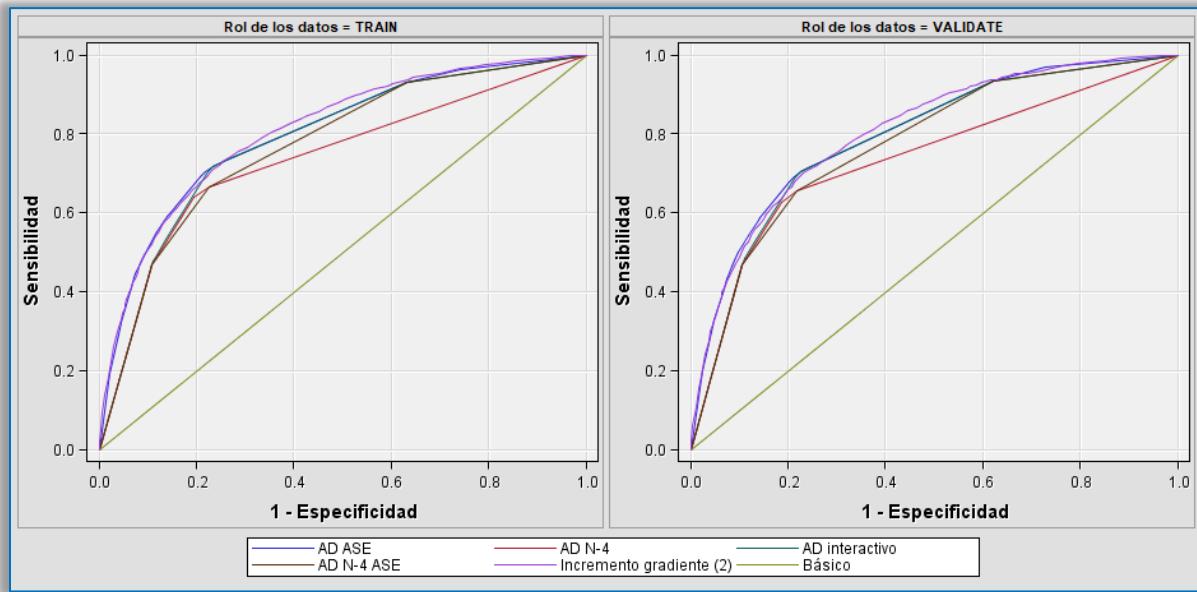
Para construir el modelo, se ha elegido 2 métodos de modelos, el árbol de decisión y modelo de regresión.

#### 3.1. Árbol de decisión

Para el método mediante el árbol de decisión se ha probado con diferentes tipos de árboles según su forma de crear los subárboles:

- Árbol de decisión interactivo: Crea el subárbol mediante el método de “evaluación”
- Árbol de decisión N-4: Crea el subárbol con número de hojas=4.
- Árbol de decisión ASE: Similar al primero, pero utiliza el error cuadrático medio como medida de evaluación.
- Árbol N-4 ASE: Similar al segundo, pero usando el error cuadrático medio como medida de evaluación.
- Incremento Gradiente: Crea árbol optimo a partir de pequeños subárboles.





Los 5 árboles presentan resultados bastantes similares.

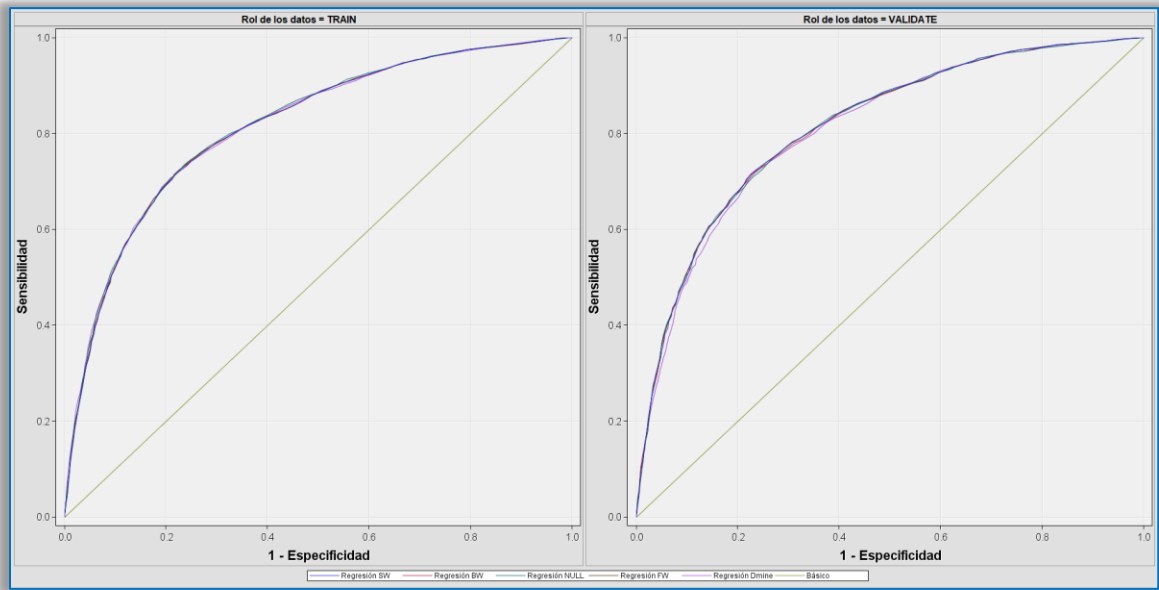
Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: Índice Roc
Y	Boost2	Incremento gradiente (2)	0.809
	Tree9	AD ASE	0.801
	Tree7	AD interactivo	0.783
	Tree10	AD N-4 ASE	0.770
	Tree8	AD N-4	0.738

Si se fija en función del AUC mediante el índice de ROC, el mejor modelo sería el que ha generado mediante el método de “incremento gradiente” con un ROC=0,809. Si se tratara de un método de árbol completo, entonces sería el método de “árbol de decisión ASE” con un ROC=0,801.

### 3.2. Modelo de regresión

Por otra parte mediante el modelo de regresión, al tratarse la variable objetiva como una variable binaria, en este caso se trabaja con modelo de regresión logística. A igual que la anterior, en este proceso también se ha generado varios modelos dependiendo de su forma de generar el modelo:

- Regresión NULL: Emplea todos los inputs para generar el modelo.
- Regresión FW: Emplea el método de selección hacia adelante.
- Regresión BW: Emplea el método de selección hacia atrás.
- Regresión SW: Emplea el método de selección paso a paso.
- Regresión Dmine: Método automático que selecciona las variables más relevantes para generar el modelo.

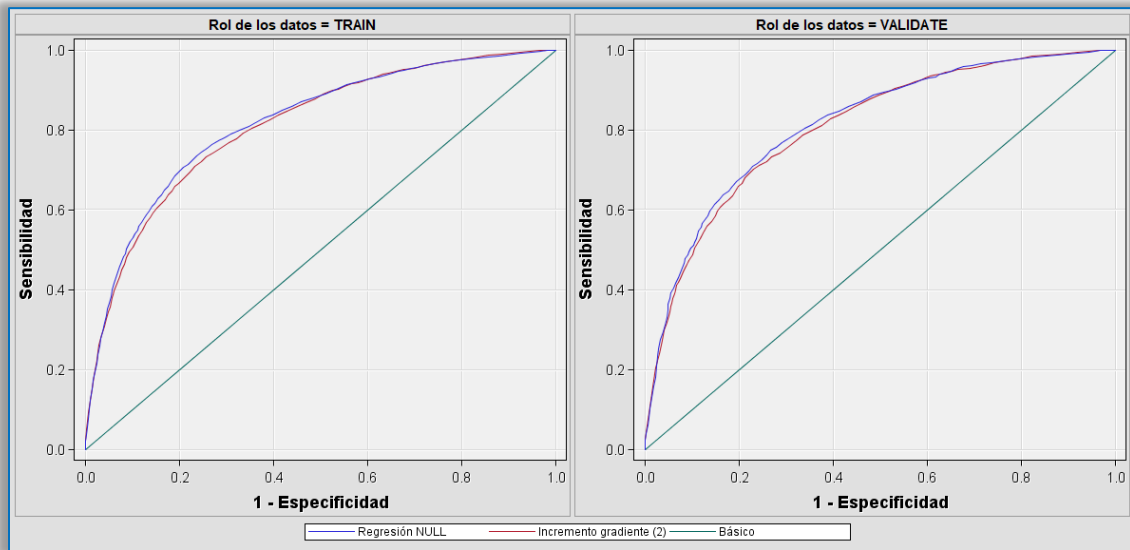


Los resultados mostrados en las gráficas también son muy similares.

Estadísticos de ajuste  
Selección de modelo basada en Entrenar: Índice Roc (\_AUR\_)

Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: Índice Roc	Entrenamiento: Error cuadrático medio
Y	Reg7	Regresión NULL	0.816	0.17472
	Reg10	Regresión FW	0.814	0.17553
	Reg8	Regresión BW	0.814	0.17553
	DmineReg	Regresión Dmine	0.814	0.17519
	Reg9	Regresión SW	0.813	0.17583

Basando en el índice ROC, el modelo de “regresión logística NULL”, es decir el modelo que ha usado todos los inputs para generar el modelo es la que presenta mejor resultado con un ROC=0,816.



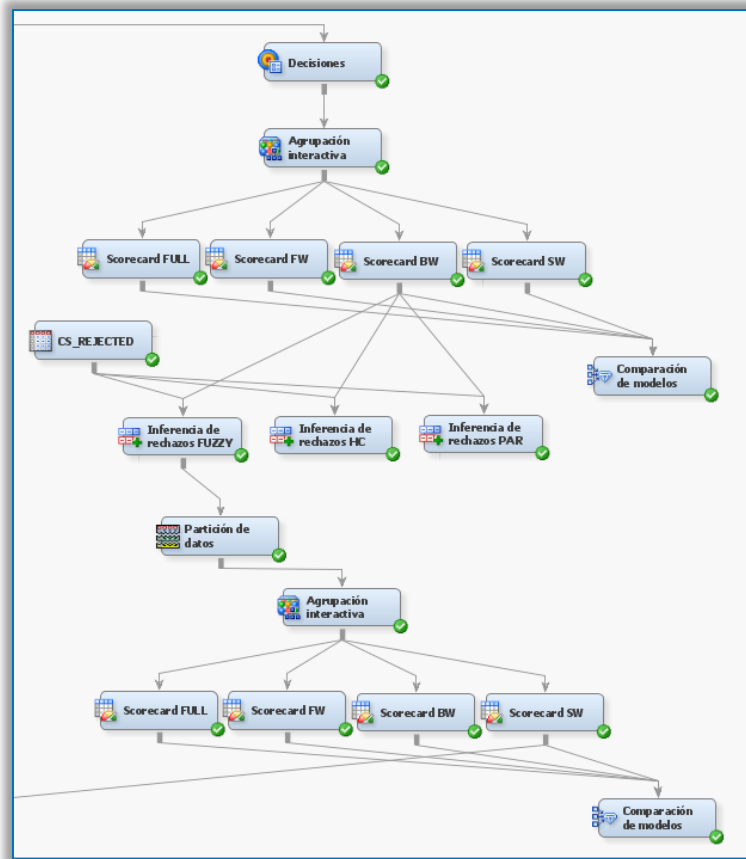
Observando las gráficas, se puede ver que el modelo de regresión obtiene mejor resultado en relación la AUC.

Estadísticos de ajuste			
Selección de modelo basada en Entrenar: Índice Roc (_AUR_)			
Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: Índice Roc
Y	Reg7	Regresión NULL	0.816
	Boost2	Incremento gradiente (2)	0.809
Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
0.17472	0.25128	0.17647	0.26261
0.17847	0.26237	0.17980	0.26568

Comparando los resultados de ROC, el modelo de “regresión NULL” muestra un ROC=0,816, mientras que el árbol de decisión generado por el “incremento gradiente” tiene un ROC=0,809. Por lo tanto se afirma de que el modelo de regresión NULL, es el modelo óptimo.

## 4. TARJETA DE PUNTUACIÓN

En el siguiente esquema se muestra todas las fases seguidas para alcanzar a la tarjeta de puntuación para el modelo.



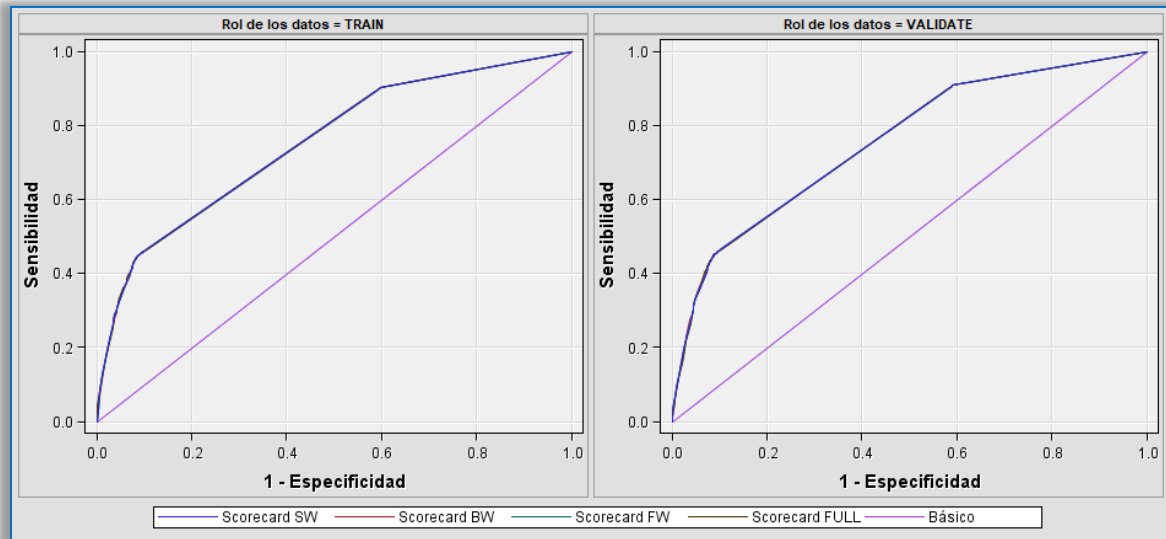
Primero se genera un nodo de “Decisiones” para usarlo al final con la salida de la puntuación. Se inserta el nodo de “agrupación Interactiva” para seleccionar las variables con mayor capacidad predictiva.

Variable	Etiqueta	Agrupamiento predefinido	Nivel	Rol calculado	Nuevo rol	Valor de información	Gini original
CB_I_presenciaCB	Indicador presencia Burea...		BINARY	Input	Predeterminado	0,708	35,901
CB_Cod_Bandas_CreditBur...	Bandas Bureau Credito		NOMINAL	Input	Predeterminado	0,687	35,131
IMP_CB_Num_CPdistintos	Imputed: CP distintos en B...		NOMINAL	Input	Predeterminado	0,679	33,485
IMP_CB_Cod_MejorSituCo...	Imputed: Mejor Situacion ...		NOMINAL	Input	Predeterminado	0,662	33,136
IMP_CB_Cod_PeorSituCon...	Imputed: Peor Situacion (...)		NOMINAL	Input	Predeterminado	0,661	33,114
IMP_REP_REP_CB_Int_Deu...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,596	28,285
IMP_REP_REP_CB_Int_Max...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,592	28,203
IMP_REP_REP_CB_Int_Deu...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,582	28,012
IMP_CB_I_DeudaTelco	Imputed: Indicador Deuda ...		BINARY	Input	Predeterminado	0,541	27,366
Neg_I_PortabilidadFijo	Indicador Portabilidad		BINARY	Input	Predeterminado	0,424	28,655
IMP_CB_I_MismaSituacion	Imputed: Indicador Misma...		BINARY	Input	Predeterminado	0,378	21,851
IMP_CB_I_EntidadFinanciera	Imputed: Tiene deuda en ...		BINARY	Input	Predeterminado	0,378	19,648
IMP_CB_I_EntidadTelco	Imputed: Indicador Cedent...		BINARY	Input	Predeterminado	0,318	17,029
IMP_CB_I_EntTelefonoMovil	Imputed: Indicador Cedent...		BINARY	Input	Rechazado	0,262	13,815

En este caso, se procede a seleccionar solamente aquellas variables con un “Índice de Valor” superior a 0,3.

Para la generación de “Scorecard” se empleó los 4 métodos:

- Scorecard FULL: Emplea todos los inputs seleccionados previamente.
- Scorecard FW: Emplea el método de selección hacia adelante.
- Scorecard BW: Emplea el método de selección hacia atrás.
- Scorecard SW: Emplea el método de selección paso a paso.



Observando la gráfica, se puede ver que los resultados son prácticamente idénticos.

Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: Índice Roc
Y	Scorecard14	Scorecard BW	0.752
	Scorecard12	Scorecard FULL	0.752
	Scorecard13	Scorecard FW	0.752
	Scorecard15	Scorecard SW	0.752
Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
0.19502	0.32001	0.19361	0.31981
0.19495	0.32008	0.19366	0.31836
0.19506	0.31980	0.19375	0.31933
0.19506	0.31980	0.19375	0.31933

Revisando el índice ROC se confirma que los 4 métodos obtienen el mismo ROC=0,752. Elegimos el primero, ya que presenta un error cuadrático medio en la “validación” más bajo.

Usando el nodo de “Scorecard BW” y relacionando con el dataset “CS\_REJECTED” se repite el mismo proceso que los pasos anteriores para comparar la eficacia de modelo con respecto a los datos de usuarios rechazados históricamente y mejorar el modelo de puntuación. Para ello se conecta estos 2 nodos, a la “inferencia de rechazos” con los 3 métodos:

- Inferencia de rechazo FUZZY: Emplea el método de inferencia Fuzzy.
- Inferencia de rechazo HC: Emplea el método de inferencia Hard Cutoff.
- Inferencia de rechazo PAR: Emplea el método de inferencia Parceling.

Inferencia de rechazos : tasas de evento			
Tasa de evento de datos de entrenamiento	Tasa de evento de datos de validación	Tasa de evento inferida de rechazos	Tasa de evento de datos aumentada
50.00	50.00	46.66	49.00

Como resultado, el método “Fuzzy” ofrece el mejor resultado de tasa de evento con respecto a los resultados de entrenamiento de validación establecida.

Variable	Etiqueta	Agrupamiento predefinido	Nivel	Rol calculado	Nuevo rol	Valor de información	Gini original
CB_I_presenciaCB	Indicador presencia Burea...		BNARY	Input	Predeterminado	0,709	35,132
IMP_CB_Num_CFdistintos	Imputed: CF distintos en B...		NOMINAL	Input	Predeterminado	0,669	32,377
CB_Cod_Bandas_CreditBur...	Bandas Bureau Credito		NOMINAL	Input	Predeterminado	0,668	33,707
IMP_CB_Cod_PeorSituCon...	Imputed: Peor Situacion (...)		NOMINAL	Input	Predeterminado	0,659	32,086
IMP_CB_Cod_MejorSituCo...	Imputed: Mejor Situacion (...)		NOMINAL	Input	Predeterminado	0,657	32,123
IMP_REP_REP_CB_Int_Deu...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,614	28,794
IMP_REP_REP_CB_Int_Deu...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,611	28,744
IMP_REP_REP_CB_Int_Max...	Imputed: Replacement: R...		INTERVAL	Input	Predeterminado	0,61	28,726
IMP_CB_I_DeudaTelco	Imputed: Indicador Deuda ...		BNARY	Input	Predeterminado	0,532	26,225
Neg_I_PortabilidadFijo	Indicador Portabilidad		BNARY	Input	Predeterminado	0,451	29,694
IMP_CB_I_MismaSituacion	Imputed: Indicador Misma...		BNARY	Input	Predeterminado	0,375	21,069
Doc_Cod_TipoDoc	Tipo de Documento		NOMINAL	Input	Rechazado	0,145	15,863

Al igual que antes, se vuelve a simplificar las variables quedando solamente con las que tienen un IV>0,3.

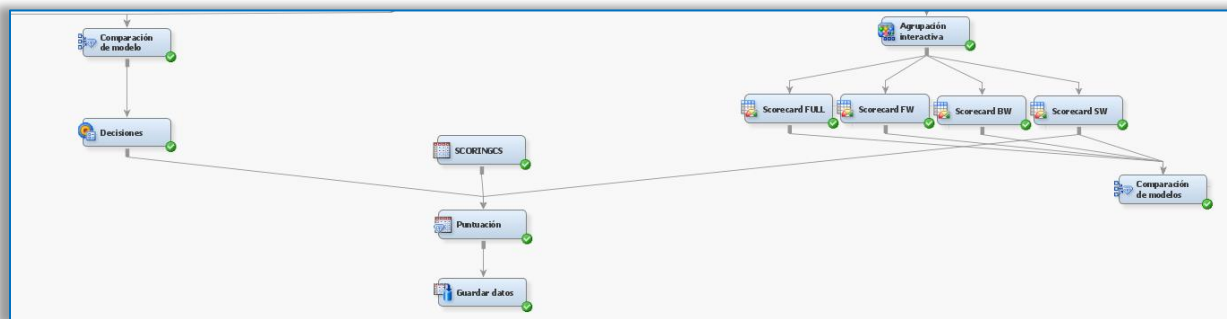
Estadísticos de ajuste			
Selección de modelo basada en Entrenar: Índice Roc (_AUR_)			
Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: Índice Roc
Y	Scorecard11	Scorecard SW	0.752
	Scorecard17	Scorecard FW	0.752
	Scorecard18	Scorecard BW	0.752
	Scorecard16	Scorecard FULL	0.752
Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
0.19435	0.31888	0.19911	0.32688
0.19435	0.31888	0.19911	0.32688
0.19435	0.31888	0.19911	0.32688
0.19433	0.31886	0.19914	0.32696

Observando los resultados devueltos por los 4 métodos de Scorecard, se puede ver que se obtiene el mismo valor de ROC=0,752. Como se ha modificado el orden de los métodos, cogeremos el primero de ellos, estableciendo ahora el “Scorecard SW” como el método óptimo para realizar la puntuación final.







Scorecard		Grupo	Puntos de scorecard	Peso de la evidencia	Tasa de eventos TARGET_I_EGINVOL_12M #1	Porcentaje de población	Coefficiente
Indicador presencia Bureau Credito	1	1.00	-54	-1.53	81.40	26.93	-0.67
	0_MISSING_ _UNKNOWN_	2.00	24	0.49	36.84	73.07	-0.67
Imputed: Indicador Deuda Telco	1	1.00	-6	-1.70	83.97	18.66	-0.11
	0_MISSING_ _UNKNOWN_	2.00	7	0.33	40.80	81.34	-0.11
Imputed: Indicador Misma Situación (Bureau Credito)	1	1.00	19	-1.53	81.40	16.13	0.17
	0_MISSING_ _UNKNOWN_	2.00	3	0.25	42.58	83.87	0.17
Imputed: Replacement: Replacement: Deuda Media (Bureau Credito)	IMP_REP_REP_CB_Int_DeudaMedia< 8961_MISSING_	1.00	9	0.37	39.86	80.00	-0.17
	8961<= IMP_REP_REP_CB_Int_DeudaMedia	2.00	-13	-1.77	84.82	20.00	-0.17
Imputed: Replacement: Replacement: Deuda Total	IMP_REP_REP_CB_Int_DeudaTotal< 8854_MISSING_	1.00	9	0.36	38.87	80.00	-0.19
	8854<= IMP_REP_REP_CB_Int_DeudaTotal	2.00	-15	-1.76	84.76	20.00	-0.19
Indicador Portabilidad	1	1.00	89	1.09	24.36	30.29	-0.88
	0_MISSING_ _UNKNOWN_	2.00	-17	-0.43	59.50	69.71	-0.88

La tabla de puntuación resultantes se muestra en esta tabla.

## 5. RESULTADOS DEL TEST CIEGO



Juntando el mejor modelo con el modelo de “regresión NULL” y el “scorecard SW” se procede a generar la puntuación para cada observación del dataset “SCORINGCS”.

	em_save_TRAIN_		11/01/2025 7:18	Hoja de cálculo d...	7.120 KB
	em_save_VALIDATE		11/01/2025 7:16	Hoja de cálculo d...	2.477 KB
	em_save_SCORE		11/01/2025 7:16	Hoja de cálculo d...	23.138 KB

Los resultados se almacenan en un fichero \*xlsx denominado “em\_save\_SCORE”, junto con los otros 2 dataset de entrenamiento “em\_save\_TRAIN” y validación “em\_save\_VALIDATE”; todos con la columna de puntuación.



U_Target_i_ECInvol_12m	b_Target_i_ECInvol_12m	EM_SEGMENT	EM_EVENTPROBABILITY	EM_PROBABILITY	EM_CLASSIFICATION
0	20	20	0,120356052	0,879643948	0
0	11	11	0,448395233	0,551604767	0
0	17	17	0,201643189	0,798356811	0
0	20	20	0,114187263	0,885812737	0
0	20	20	0,107276342	0,892723658	0
0	20	20	0,111828372	0,888171628	0
0	17	17	0,17299125	0,82700875	0
0	17	17	0,207038427	0,792961573	0
0	20	20	0,120080747	0,879919253	0
1	5	5	0,780072137	0,780072137	1
0	19	19	0,125698403	0,874301597	0
0	20	20	0,109486011	0,890513989	0
0	20	20	0,107276342	0,892723658	0
1	7	7	0,663500135	0,663500135	1
0	16	16	0,233550467	0,766449533	0
0	18	18	0,149041617	0,850958383	0
0	20	20	0,121493781	0,878506219	0
1	9	9	0,545623805	0,545623805	1
0	18	18	0,149041617	0,850958383	0
0	12	12	0,397031914	0,602968086	0
0	10	10	0,494671915	0,505328085	0
0	19	19	0,125523364	0,874476636	0
1	8	8	0,576452514	0,576452514	1
0	20	20	0,093000371	0,906999629	0
1	6	6	0,701430119	0,701430119	1
0	17	17	0,17362009	0,82637991	0
0	17	17	0,175390903	0,824609097	0
1	6	6	0,729199985	0,729199985	1
0	20	20	0,099671205	0,900328795	0
0	16	16	0,241389292	0,758610708	0

La columna de puntuación denominado “EM\_CLASSIFICATION” muestra los resultados en 0 y 1; en la cual el valor “0” se refiere a un usuario con una alta probabilidad de pagar, con lo cual se le puede aprobar la solicitud, mientras que el valor “1” se refiere a aquellos usuarios que tiene un alto riesgo de impago, por lo tanto se le denegaría la solicitud.