

Ejercicio 1

Conecta las salidas, ejecuta el modelo y analiza las características y estadísticas de los atributos de los datasets

a. Tipos de datos

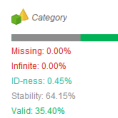
- Polinomiales
- Id y Edad numéricos

b. Outliers y faltantes

- No hay datos faltantes
- No hay outliers en Edad

< > ActividadWebsite

Summary



Top Values

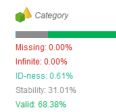


3 Distinct Values:

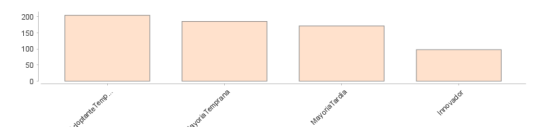
Value	Count	Percentage
Escasa	424	64.15%
Regular	183	27.69%
Frecuente	54	8.17%

< > AdopcionEReader

Summary



Top Values

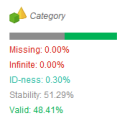


4 Distinct Values:

Value	Count	Percentage
AdoptanteTemprano	205	31.01%
MayoriaTemprana	186	28.14%
MayoriaTardia	172	26.02%
Innovador	98	14.83%

< > ComproElectronicos12

Summary



Top Values

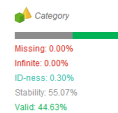


2 Distinct Values:

Value	Count	Percentage
Si	339	51.29%
No	322	48.71%

< > ComproLibrosDigitales

Summary



Top Values

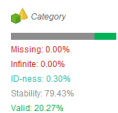


2 Distinct Values:

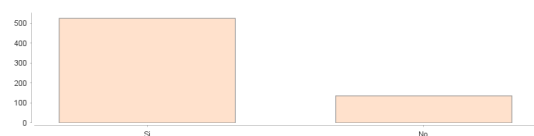
Value	Count	Percentage
No	364	55.07%
Si	297	44.93%

< > ComproMedios18

Summary



Top Values

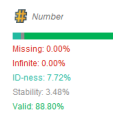


2 Distinct Values:

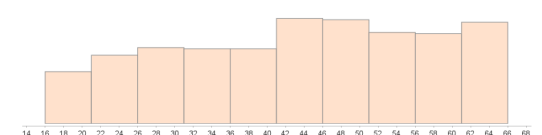
Value	Count	Percentage
Si	525	79.43%
No	136	20.57%

< > Edad

Summary



Distribution

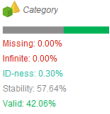


Statistics

Name	Value
Minimum	16
Maximum	66
Average	42.794
Standard Deviation	13.862

< > EstadoCivil

Summary



Top Values

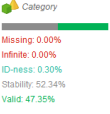


2 Distinct Values:

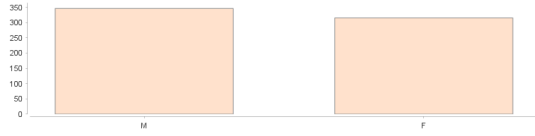
Value	Count	Percentage
C	381	57.64%
S	280	42.36%

< > Sexo

Summary



Top Values

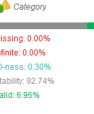


2 Distinct Values:

Value	Count	Percentage
M	346	52.34%
F	315	47.66%

< > MiroElectronicos12

Summary



Top Values



2 Distinct Values:

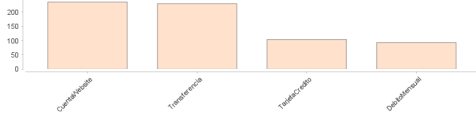
Value	Count	Percentage
Si	613	92.74%
No	48	7.26%

< > MetodoPago

Summary



Top Values



4 Distinct Values:

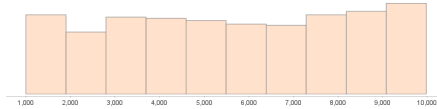
Value	Count	Percentage
CuentaWebSite	235	35.55%
Transferencia	229	34.64%
TarjetaCredito	104	15.73%
DebitoMensual	93	14.07%

< > ID

Summary



Distribution



Statistics

Name	Value
Minimum	1003
Maximum	9982
Average	5638.188
Standard Deviation	2635.213

ID se debe filtrar ya que no aporta información valiosa al problema, por lo que su presencia solamente representa ruido que puede desembocar sobre ajuste sobre atributos irrelevantes.

Se setea el rol de id al atributo **ID** y label al atributo **AdopcionEReader**.

Ejercicio 2

- 1) Ejecutar el modelo y observar el árbol y los resultados (el árbol no ha cambiado todavía).

No se adjunta el resultado ya que el árbol resultante (tanto en su representación visual como en formato de reglas es demasiado extenso).

Es claro que pueden (y deben) realizarse optimizaciones para lograr un árbol de decisión manejable.

- 2) Pasar a la vista de estadísticas y estudiar los valores obtenidos para la predicción y los niveles de confianza

Row No.	ID	prediction(A...	confidence(...	confidence(L...	confidence(...	confidence(...	Sexo	Edad	EstadoC
1	56031	Innovador	0.030	0.576	0.318	0.076	M	57	S
2	25913	AdoptanteTe...	0.047	0.203	0.547	0.203	F	51	C
3	19396	MayoriaTardia	0.846	0.038	0.038	0.077	M	41	C
4	93666	MayoriaTemp...	0.250	0	0	0.750	M	66	S
5	72282	MayoriaTardia	0.846	0.038	0.038	0.077	F	31	S
6	64466	MayoriaTemp...	0.250	0	0	0.750	M	68	C
7	76655	MayoriaTardia	0.842	0.018	0.061	0.079	F	51	S
8	48465	MayoriaTardia	0.500	0.500	0	0	F	36	S
9	19889	AdoptanteTe...	0.047	0.203	0.547	0.203	M	29	C
10	63570	MayoriaTemp...	0	0	0	1	M	61	C
11	63239	AdoptanteTe...	0.047	0.203	0.547	0.203	M	47	S
12	67603	MayoriaTemp...	0	0	0	1	F	62	S

- 3) Cambiar el criterio de división a “Gini Index” y estudiar el árbol generado, así como los resultados obtenidos. Se puede observar que el árbol es bastante más detallado que el anterior

El árbol resultante es excesivamente grande, tiene una profundidad de 10 niveles y es considerablemente más detallado y acoplado a los datos de entrenamiento que el árbol con las configuraciones default.

- 4) Experimentar variando los valores de los parámetros de máxima profundidad, cantidad de elementos para dividir un nodo y máxima cantidad de elementos en las hojas, y registrar los diversos resultados en la tabla.

Asignando la profundidad máxima de 5:

Row No.	ID	prediction(A...	confidence(...	confidence(...	confidence(...	confidence(...	Sexo	Edad	EstadoC
1	56031	Innovador	0.059	0.529	0.294	0.118	M	57	S
2	25913	MayoriaTemp...	0.098	0.098	0.235	0.569	F	51	C
3	19396	MayoriaTardia	0.751	0.021	0.053	0.175	M	41	C
4	93666	Innovador	0	0.636	0.182	0.182	M	66	S
5	72282	MayoriaTardia	0.751	0.021	0.053	0.175	F	31	S
6	64466	MayoriaTemp...	0.098	0.098	0.235	0.569	M	68	C
7	76655	MayoriaTardia	0.751	0.021	0.053	0.175	F	51	S
8	48465	Innovador	0.125	0.875	0	0	F	36	S
9	19889	AdoptanteTe...	0	0.083	0.875	0.042	M	29	C
10	63570	MayoriaTemp...	0	0	0.333	0.667	M	61	C
11	63239	MayoriaTemp...	0.098	0.098	0.235	0.569	M	47	S
12	67603	MayoriaTemp...	0.098	0.098	0.235	0.569	F	62	S
13	65685	AdoptanteTe...	0.125	0.250	0.562	0.062	M	32	C
14	77373	MayoriaTemp...	0.083	0.028	0	0.889	M	17	C

Asignando el min size for split en 8:

Row No.	ID	prediction(A...	confidence(...	confidence(...	confidence(...	confidence(...	Sexo	Edad	EstadoC
1	56031	AdoptanteTe...	0	0.286	0.429	0.286	M	57	S
2	25913	AdoptanteTe...	0.200	0.100	0.600	0.100	F	51	C
3	19396	MayoriaTardia	0.846	0.038	0.038	0.077	M	41	C
4	93666	Innovador	0	0.636	0.182	0.182	M	66	S
5	72282	MayoriaTardia	0.846	0.038	0.038	0.077	F	31	S
6	64466	MayoriaTardia	0.500	0.167	0	0.333	M	68	C
7	76655	MayoriaTardia	0.842	0.018	0.061	0.079	F	51	S
8	48465	Innovador	0.125	0.875	0	0	F	36	S
9	19889	AdoptanteTe...	0	0.067	0.867	0.067	M	29	C
10	63570	MayoriaTemp...	0	0	0	1	M	61	C
11	63239	MayoriaTemp...	0	0.167	0	0.833	M	47	S
12	67603	MayoriaTemp...	0	0.067	0	0.933	F	62	S
13	65685	AdoptanteTe...	0.250	0	0.750	0	M	32	C
14	77373	MayoriaTemp...	0	0	0	1	M	17	C

Asignando el tamaño de hoja mínimo en 10:

Row No.	ID	prediction(A...	confidence(...	confidence(...	confidence(...	confidence(...	Sexo	Edad	EstadoC
1	56031	AdoptanteTe...	0.038	0.308	0.538	0.115	M	57	S
2	25913	AdoptanteTe...	0.100	0.100	0.600	0.200	F	51	C
3	19396	MayoriaTardia	0.846	0.038	0.038	0.077	M	41	C
4	93666	AdoptanteTe...	0.038	0.308	0.538	0.115	M	66	S
5	72282	MayoriaTardia	0.846	0.038	0.038	0.077	F	31	S
6	64466	MayoriaTemp...	0.143	0.095	0	0.762	M	68	C
7	76655	MayoriaTardia	0.842	0.018	0.061	0.079	F	51	S
8	48465	Innovador	0.167	0.417	0.333	0.083	F	36	S
9	19889	AdoptanteTe...	0.105	0.053	0.842	0	M	29	C
10	63570	Innovador	0.167	0.417	0.333	0.083	M	61	C
11	63239	MayoriaTemp...	0.050	0.100	0.300	0.550	M	47	S
12	67603	MayoriaTemp...	0.143	0.095	0	0.762	F	62	S
13	65685	AdoptanteTe...	0.105	0.053	0.842	0	M	32	C
14	77373	MayoriaTemp...	0.083	0.028	0	0.889	M	17	C

Se puede observar que la configuración que arroja el árbol menos complejo y acoplado a los datos de entrenamiento es el que cuenta con profundidad máxima de 5 y criterio de división Gini.