

Análise de Inadimplência

Estudo de Caso

Objetivos

Análise da Base de Dados

Algoritmo de Redução da Inadimplência

Conclusões

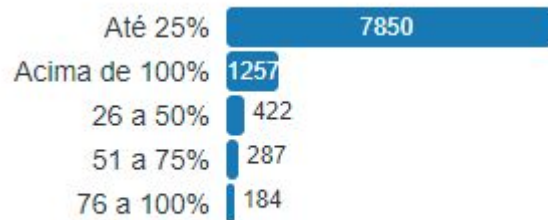
Análise da Base de Dados

#	Atributos		Descrição
1	N° do Cliente		Número interno de identificação do cliente
2	Safra		Data de aprovação do cliente
3	Estado		Estado de nascimento do cliente
4	Renda Mensal		Renda mensal estimada
5	Endividamento		Faixa de endividamento em relação ao salário mensal
6	Quantidade de Cheques sem Fundo		
7	Quantidade de Restritivos		
8	Valor dos Restritivos		
9	Quantidade de Protestos		
10	Valor dos Protestos		
11	Modelo Score 1	Pontuação do modelo (quanto maior melhor, ou seja, menor risco de crédito)	
12	Modelo Score 2	Pontuação do modelo (quanto maior melhor, ou seja, menor risco de crédito)	
13	Modelo Score 3	Pontuação do modelo (quanto maior melhor, ou seja, menor risco de crédito)	
14	Modelo Score 4	Pontuação do modelo (quanto maior melhor, ou seja, menor risco de crédito)	
15	Performance 30D3M EVER	MAU = cliente apresentou atraso > 30 dias em 3 meses de observação	
16	Performance 60D6M EVER	MAU = cliente apresentou atraso > 60 dias em 6 meses de observação	
17	Performance 60D9M EVER	MAU = cliente apresentou atraso > 60 dias em 9 meses de observação	
18	Performance 90D9M EVER	MAU = cliente apresentou atraso > 90 dias em 9 meses de observação	
19	Performance 90D12M EVER	MAU = cliente apresentou atraso > 90 dias em 12 meses de observação	
-	30D3M	Normalização do atributo "Performance 30D3M EVER"	
-	60D6M	Normalização do atributo "Performance 60D6M EVER"	
-	60D9M	Normalização do atributo "Performance 60D9M EVER"	
-	90D9M	Normalização do atributo "Performance 90D9M EVER"	
-	90D12M	Normalização do atributo "Performance 90D12M EVER"	
-	'% Endividamento'	Normalização do atributo "Endividamento"	

Análise da Base de Dados

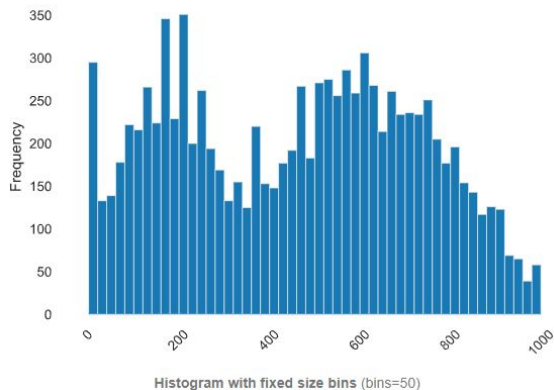
Value	Count	Frequency (%)
rs	1377	13.8%
sp	1289	12.9%
ms	1209	12.1%
ba	1073	10.7%
mg	1009	10.1%
pr	921	9.2%
go	841	8.4%
sc	807	8.1%
rj	753	7.5%
to	721	7.2%

Endividamento

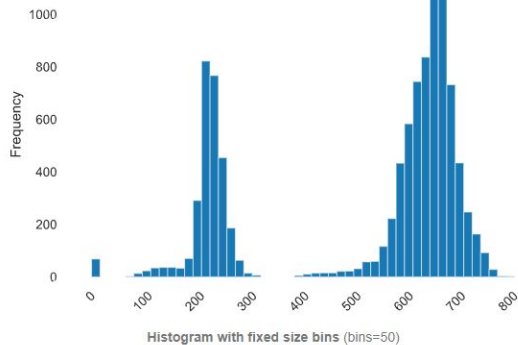


Análise da Base de Dados

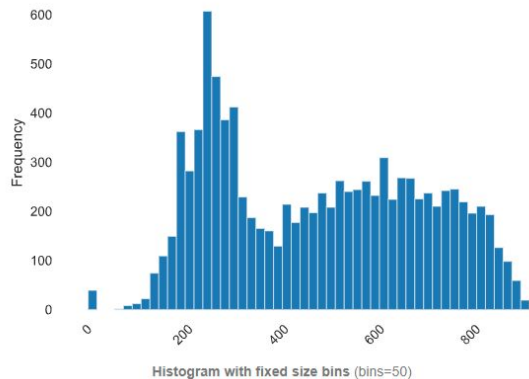
Modelo Score 1



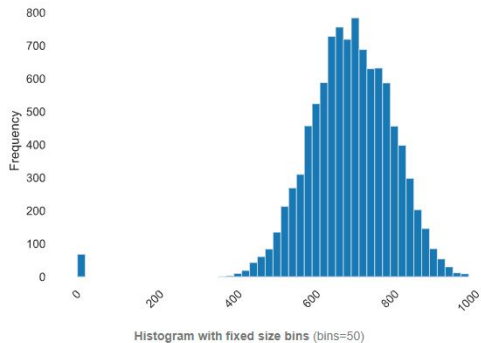
Modelo Score 2



Modelo Score 3



Modelo Score 4

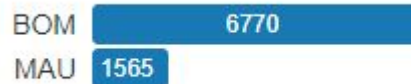


Análise da Base de Dados

Performance 30D3M EVER



Performance 90D9M EVER



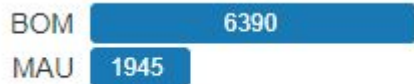
Performance 60D6M EVER



Performance 90D12M EVER



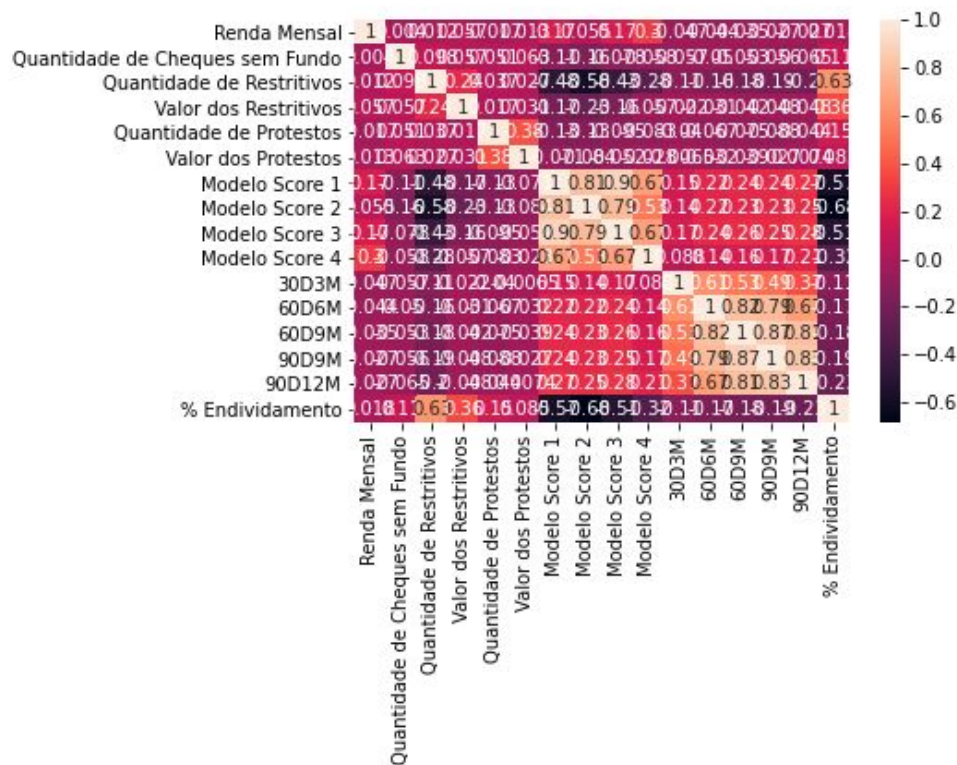
Performance 60D9M EVER



Análise da Base de Dados

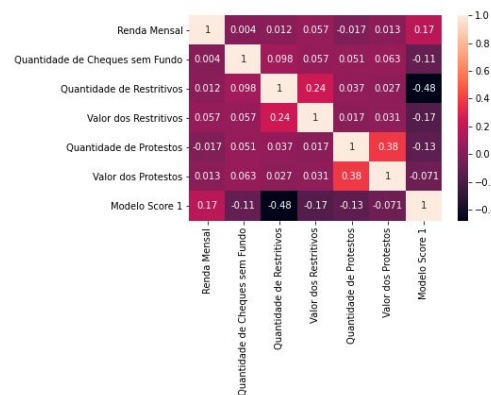
	Nº do cliente	Renda Mensal	Quantidade de Cheques sem Fundo	Quantidade de Restritivos	Valor dos Restritivos	Quantidade de Protestos	Valor dos Protestos	Modelo Score 1	Modelo Score 2	Modelo Score 3	Modelo Score 4	30D3M	60D6M	60D9M	90D9M	90D12M	% Endividamento
count	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	8335.000000	8335.000000	3354.000000	10000.000000
mean	6.101154e+08	2780.155700	0.164900	0.711100	1635.66340	0.152900	209.724300	459.182800	524.183800	483.280100	695.888500	0.875600	0.829000	0.766647	0.812238	0.762671	0.445825
std	3.515864e+08	1464.592111	1.564146	1.823285	10741.23154	1.173908	2700.805532	264.013507	200.494433	215.561897	115.237171	0.330054	0.376528	0.422991	0.390546	0.425509	0.422184
min	1.241258e+06	1041.000000	0.000000	-7.000000	0.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000
25%	3.056783e+08	1404.000000	0.000000	0.000000	0.000000	0.000000	0.000000	216.000000	253.000000	277.000000	630.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.250000
50%	6.101154e+08	3365.000000	0.000000	0.000000	0.000000	0.000000	0.000000	481.000000	627.000000	482.000000	700.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.250000
75%	9.145524e+08	3492.000000	0.000000	1.000000	145.000000	0.000000	0.000000	672.250000	669.000000	669.000000	772.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.250000
max	1.218989e+09	33564.000000	64.000000	35.000000	462266.000000	37.000000	133742.000000	999.000000	811.000000	922.000000	999.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.500000

Análise da Base de Dados

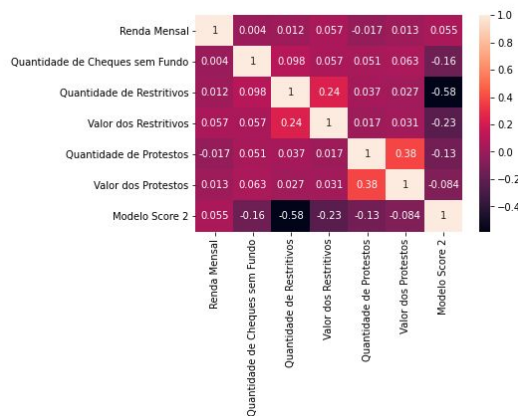


Análise da Base de Dados

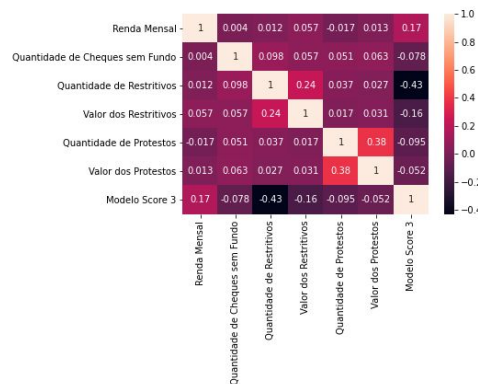
Modelo Score 1



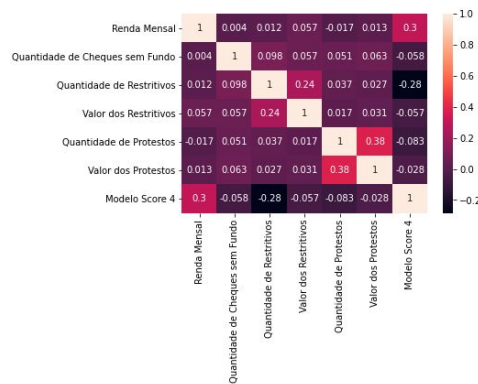
Modelo Score 2



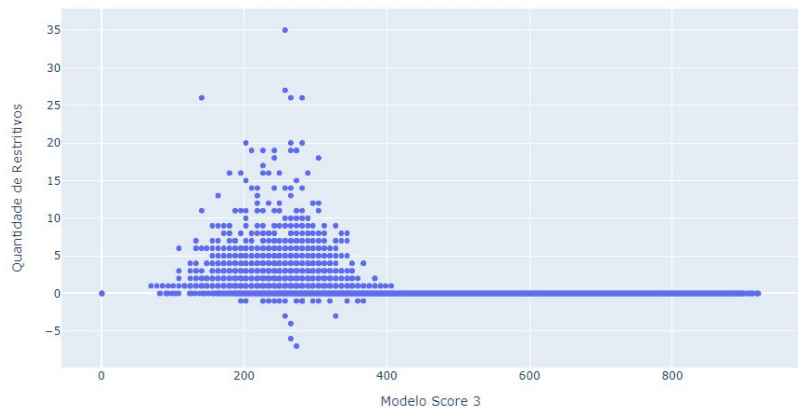
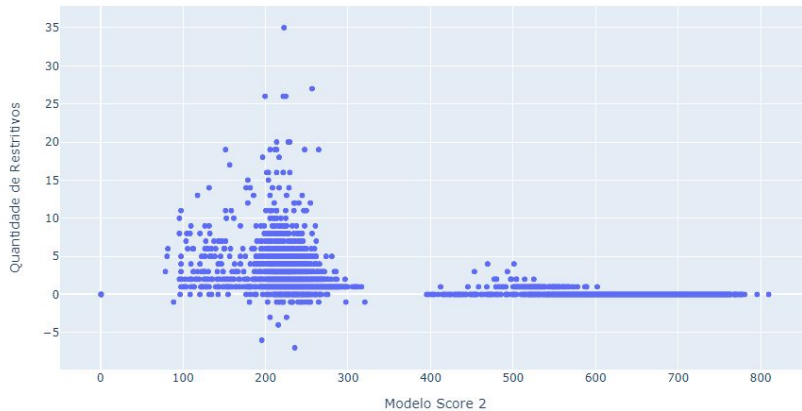
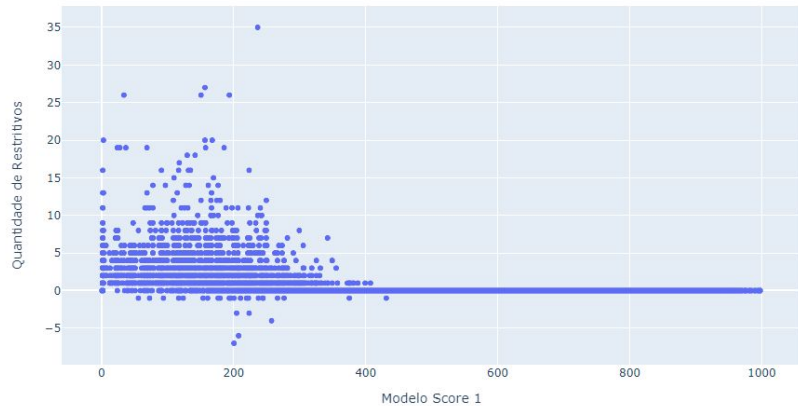
Modelo Score 3



Modelo Score 4



Análise da Base de Dados



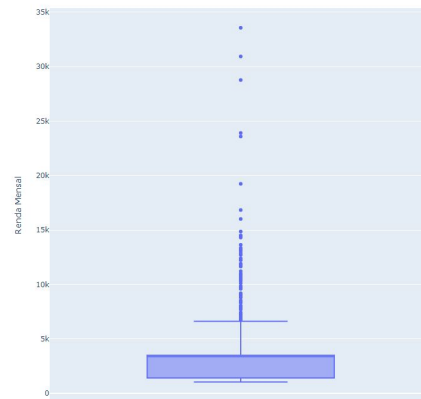
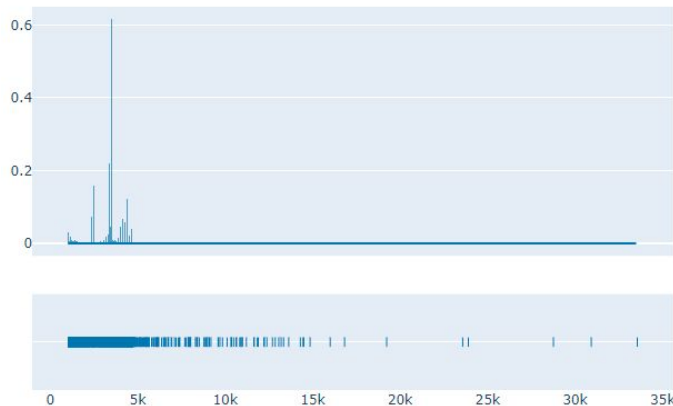
Análise da Base de Dados

Analizando Outliers “Renda Mensal”

```
count    10000.000000
mean      2780.155700
std       1464.592111
min       1041.000000
25%       1404.000000
50%       3365.000000
75%       3492.000000
max       33564.000000
Name: Renda Mensal, dtype: float64
```

Coeficiente de Pearson:
0.8518454055606476

Valor entre -1 e 1 - distribuição simétrica.
Valor maior que 1 - distribuição assimétrica positiva.
Valor maior que -1 - distribuição assimétrica negativa.



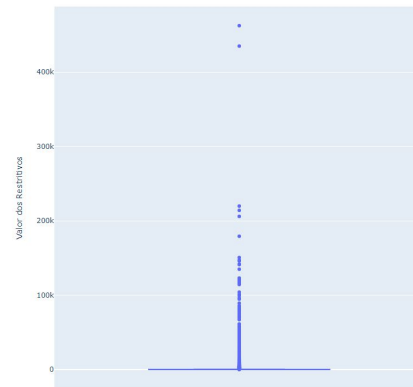
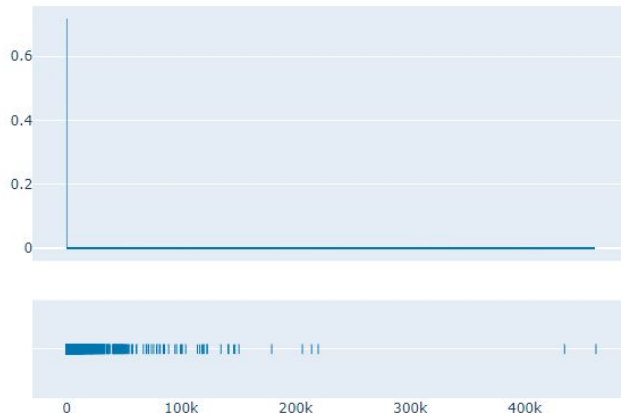
Análise da Base de Dados

Analisando Outliers “Valor dos Restritivos”

```
count    10000.00000  
mean      1635.66340  
std       10741.23154  
min        0.00000  
25%        0.00000  
50%        0.00000  
75%       145.00000  
max      462266.00000  
Name: Valor dos Restritivos, dtype: float64
```

Coeficiente de Pearson:
21.080108243066228

Valor entre -1 e 1 - distribuição simétrica.
Valor maior que 1 - distribuição assimétrica positiva.
Valor maior que -1 - distribuição assimétrica negativa.



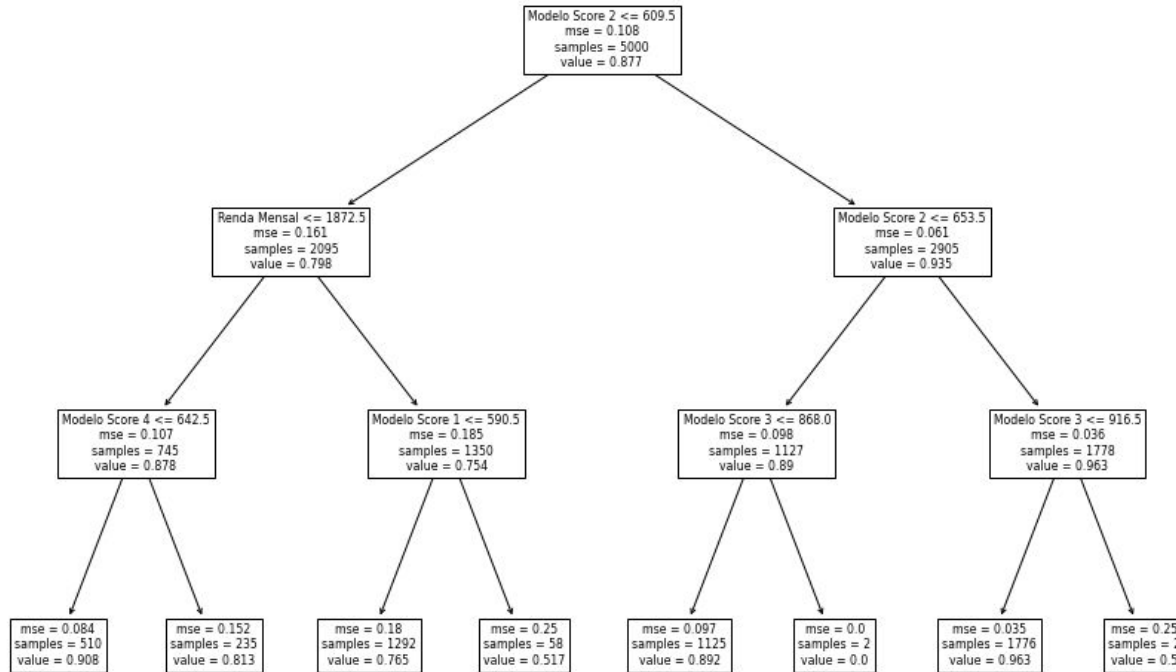
Algoritmo de Redução da Inadimplência

Árvore de Decisão

Parâmetro: 30D3M

Profundidade: Nível 3

Erro: 0.3209063038510057



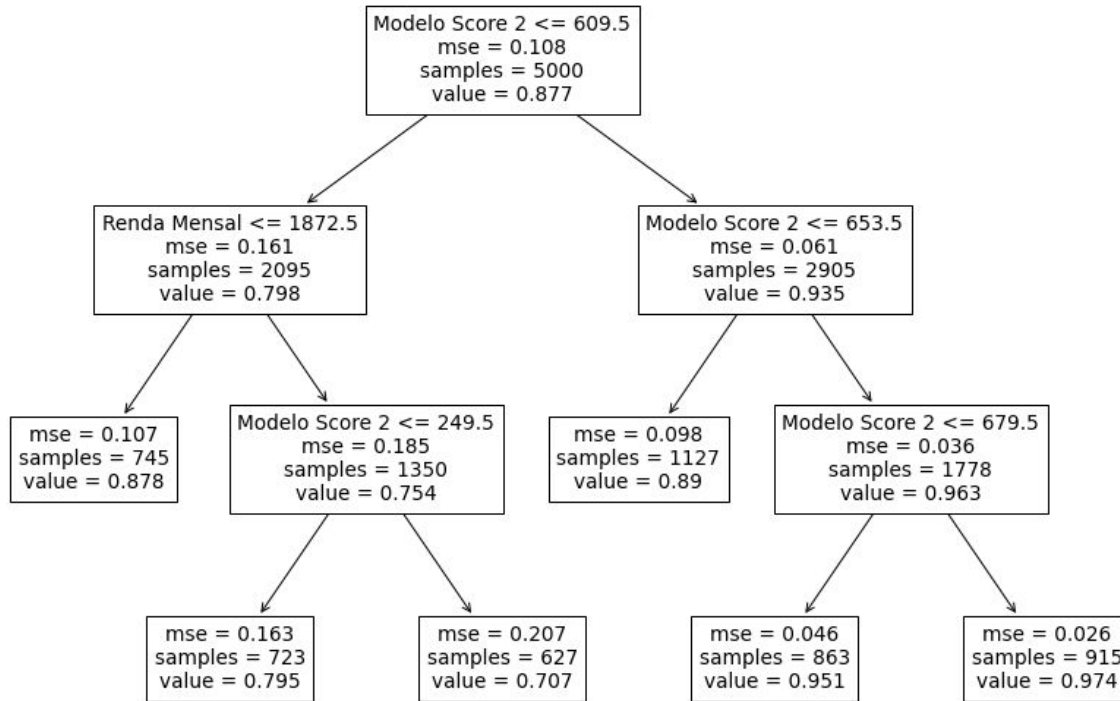
Algoritmo de Redução da Inadimplência

Árvore de Decisão

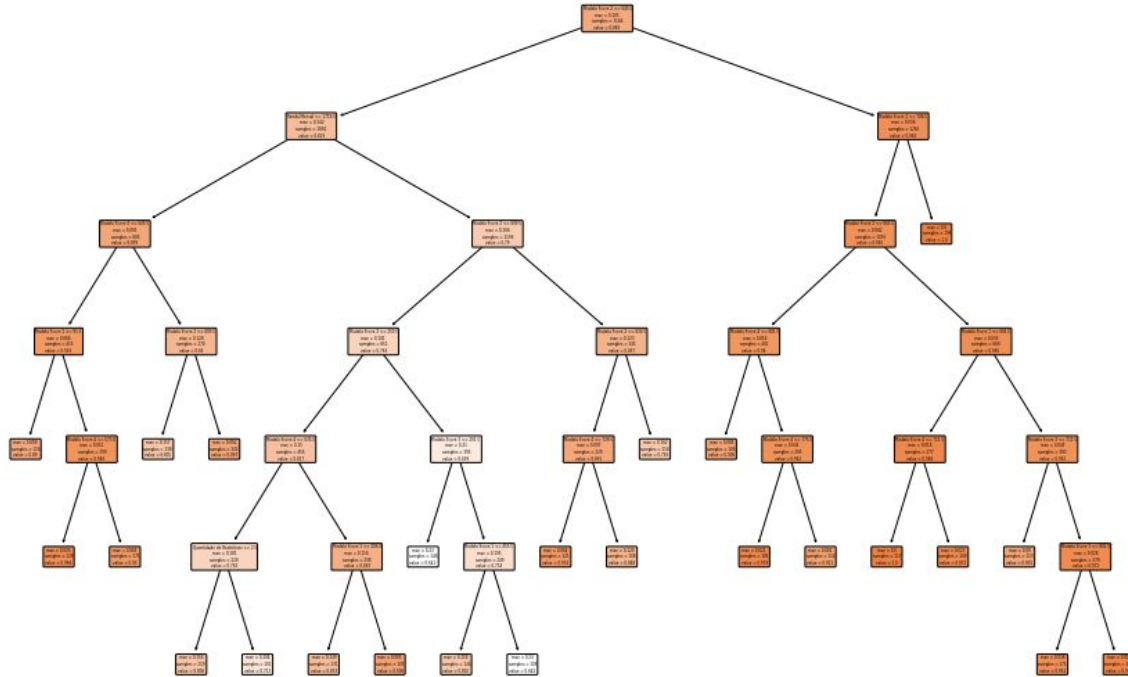
Parâmetro: 30D3M

MSL: 610

Erro: 0.3205826152987339



Algoritmo de Redução da Inadimplência



Árvore de Decisão Randômica

Parâmetro: 30D3M

N ESTIMATORS: 1000

MSL: 100

Erro: 0.3176977385719225

Conclusões

- Necessário realizar a limpeza de Outliers dos atributos de **“Renda Mensal”** e **“Valor dos Restritivos”**
- Existe uma forte correlação entre os valores de Score dos modelos com a **“Quantidade de Restritivos”**
- O **“Modelo Score 4”** apresentou a pior distribuição relativa a **“Quantidade de Restritivos”**
- O **“Modelo Score 2”** teve a melhor performance nos algoritmos utilizados para o parâmetro normalizado **“30D3M”**
- Entre os algoritmos utilizados o com menor índice de erro foi a árvore de decisão randômica.