Previsão do Score

HONG KONG

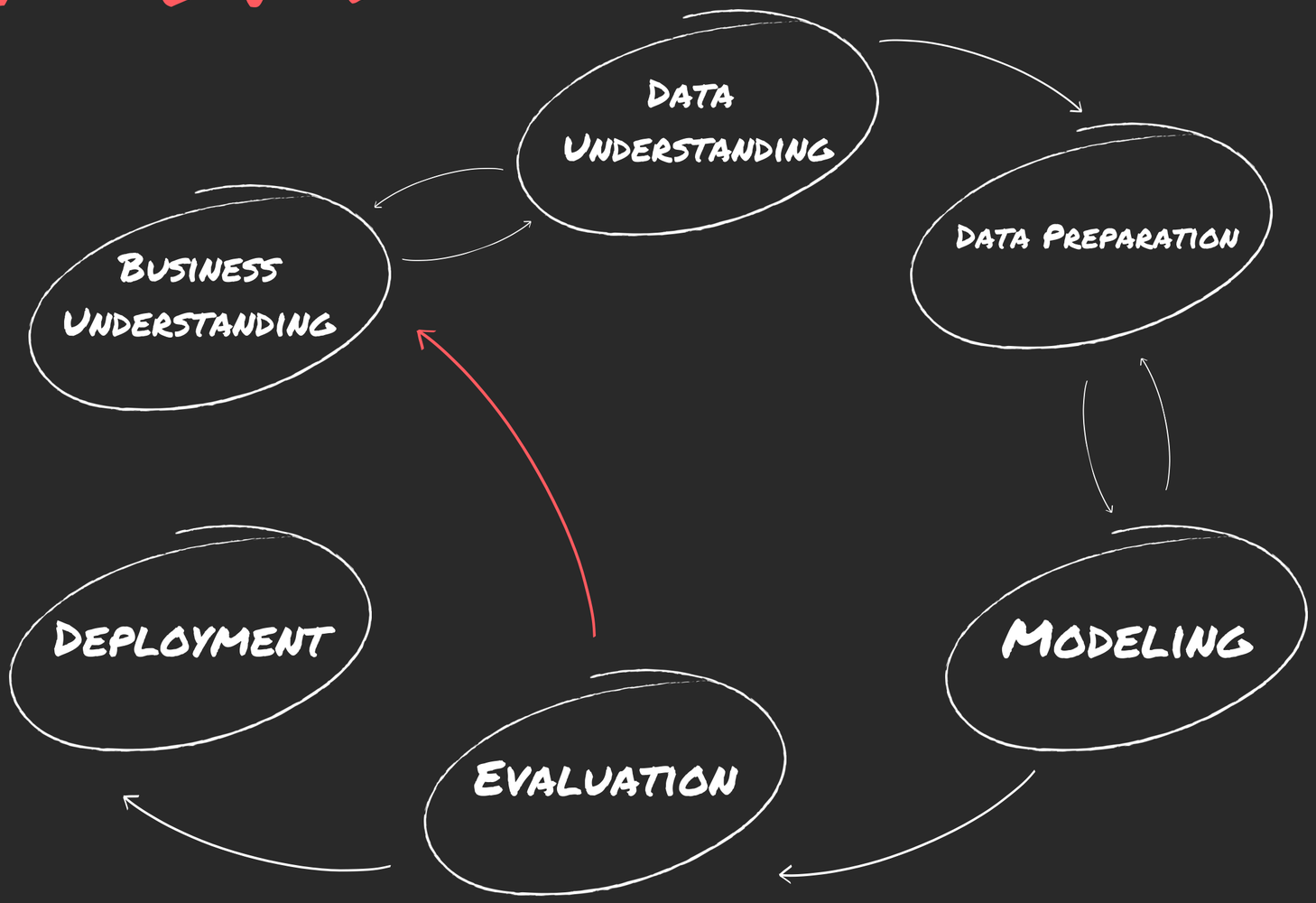
Equipa

Guilherme Mendonça

Marta Neves

René Porto

CRISP-DM


Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Business Understanding

04

Business understanding

- Objetivo do Airbnb
- Avaliação detalhada
- Objetivo técnico
- Plano do projeto

04

Data Understanding

05

Criação dos URL's

Procura de alojamentos

Extração de variáveis

06

Data Preparation

- 17.112 acomodações
- Valores omissos
- Tipo de acomodação
- Outliers
- Correlação

06

Modeling

Regressão Linear - Múltipla

AD - Bagging

Rede Neuronal

Regressão Linear - Múltipla (VC)

AD - Boosting

Árvore de Decisão

AD - Florestas Aleatórias

07

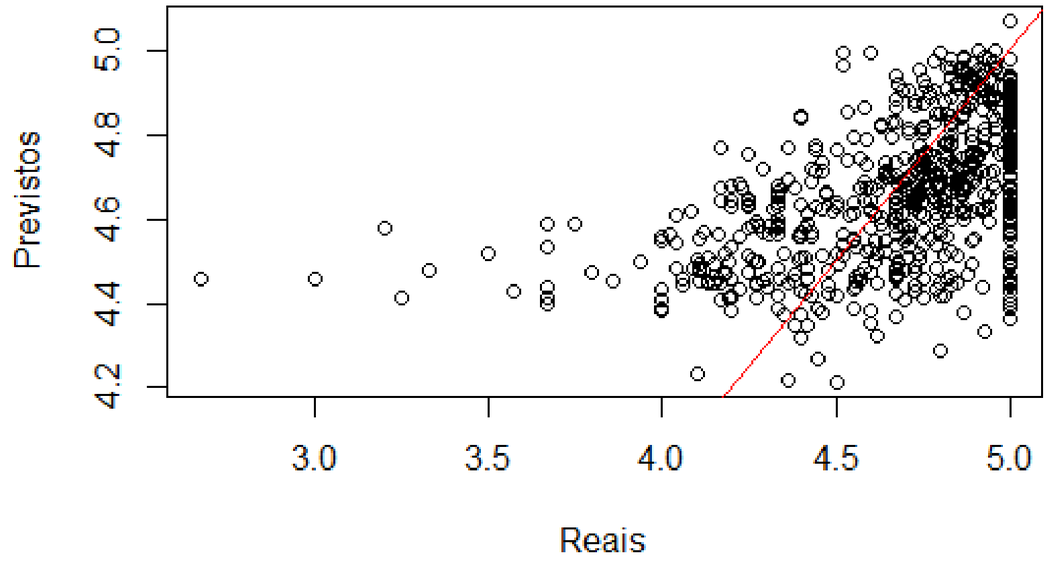
Regressão Linear

0,286

Redução de variáveis (p.value) > 0.05

Treino (70%) e Teste (30%)

Validação Cruzada

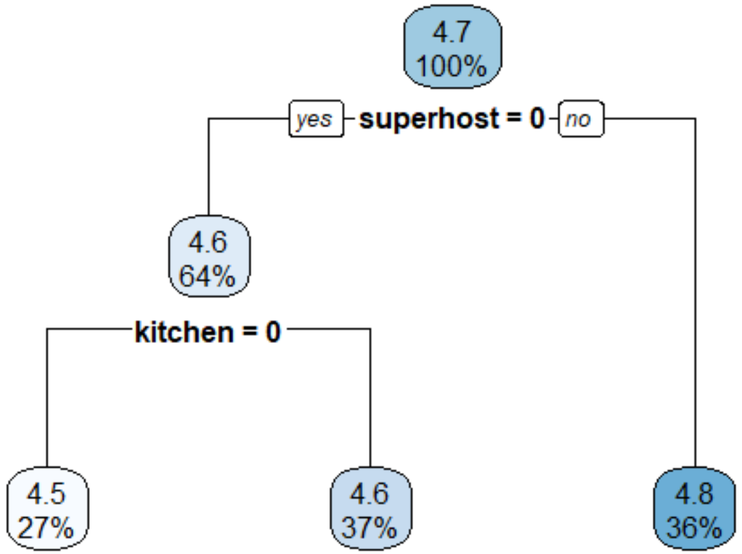


08

Árvore de Decisão

0,286

Poda (cp = 0.025)



09

Árvore de Decisão

nbagg	RMSE	R-Square
100	0.2932	0.2526
200	0.2938	0.245
300	0.2937	0.2503
500	0.2932	0.2507

Árvore de Decisão

		RMSE	R-square
1º	Depth=3 n.trees=1000 Shrinkage=0.02 n.minobsinnode= 20	0.2927	0.2535
2º	Depth=2 n.trees=500 Shrinkage=0.05 n.minobsinnode= 15	0.2919	0.2564
3º	Depth=2 n.trees=500 Shrinkage=0.05 n.minobsinnode= 5	0.2923	0.2529snii
4º	Depth=2 n.trees=300 Shrinkage=0.05 n.minobsinnode= 5	0.2894	0.2707
5º	Depth=2 n.trees=1000 Shrinkage=0.02 n.minobsinnode= 20	**0.2874**	**0.2749**

Árvore de Decisão

FLORESTAS ALEATÓRIAS

0,288

mtry	splitrule	RMSE	Rsquared	MAE
2	variance	0.2882932	0.2886527	0.2042981
2	extratrees	0.2905099	0.2788480	0.2057411
5	variance	0.2877067	0.2753290	0.2019657
5	extratrees	0.2877417	0.2747828	0.2016227
9	variance	0.2966053	0.2445166	0.2076130
9	extratrees	0.2940619	0.2549461	0.2051103
13	variance	0.3017568	0.2290373	0.2106326
13	extratrees	0.2984459	0.2431941	0.2078306
17	variance	0.3044577	0.2220346	0.2126835
17	extratrees	0.3016974	0.2341544	0.2099348

12

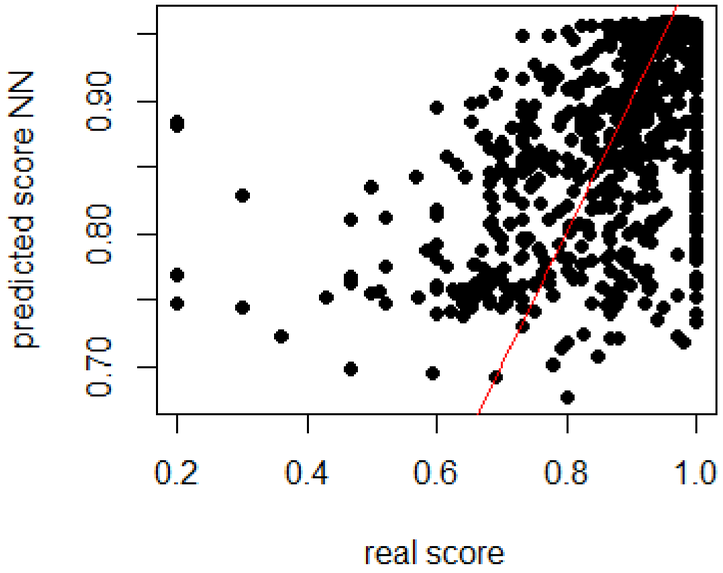
Redes Neuronais

0,117

Redução de variáveis

Normalização

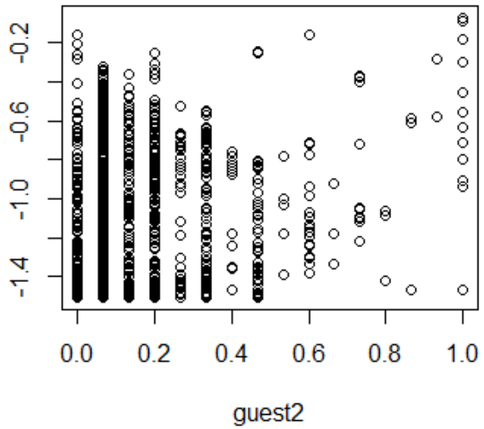
Treino (70%) e Teste (30%)

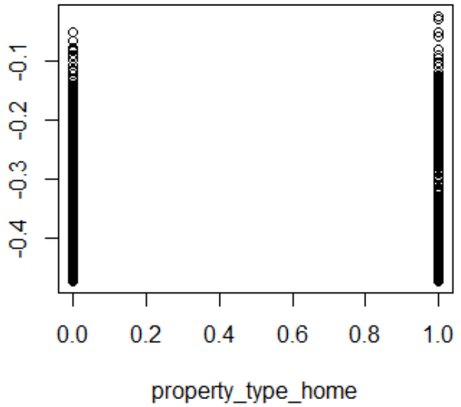


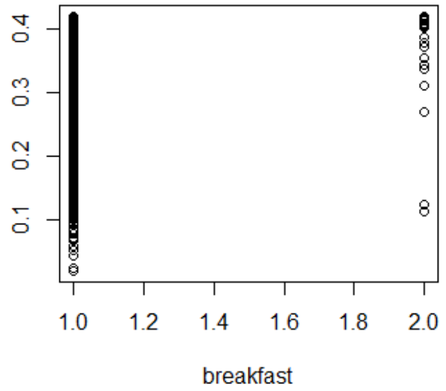
Tentativa	%Treino	%Teste	Hidden	Algoritmo	Tx.Aprendizagem	RMSE	MAPE
1	70	30	1,20	backprop	0.001	0.1179	0.1245
2	**70**	**30**	**1,30**	**backprop**	**0.001**	**0.1178**	**0.1245**
3	70	30	1,50	backprop	0.001	0.1179	0.1245
4	70	30	10,10	backprop	0.001	0.1331	0.1286
5	70	30	20,20	backprop	0.001	0.1362	0.1306

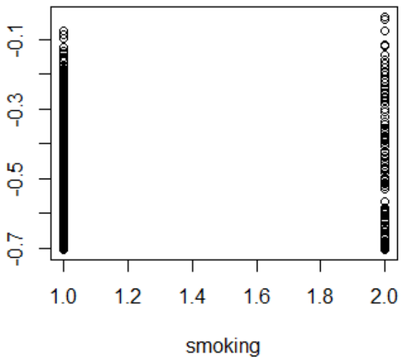
Redes Neuronais

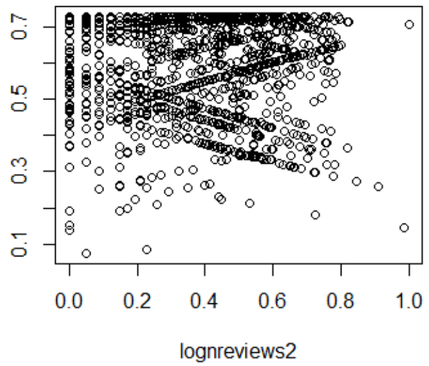
IMPORTÂNCIA DAS VARIÁVEIS

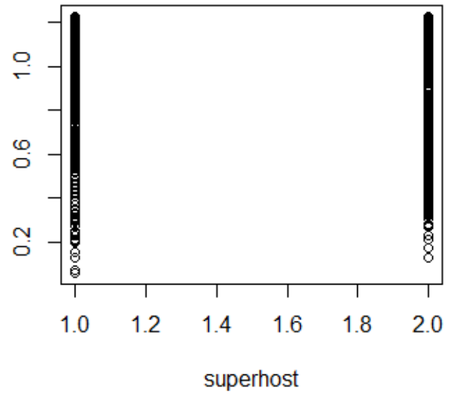












Evaluation

Regressão Linear - Múltipla → 0,286

AD - Bagging → 0,288

Regressão Linear - Múltipla (VC) → 0,286

AD - Boosting → 0,287

Árvore de Decisão → 0,290

AD - Florestas Aleatórias → 0,288

Rede Neuronal → 0,117

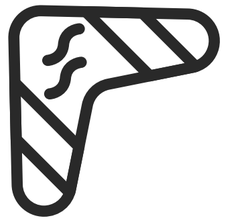
Deployment

Clique para prever o futuro!

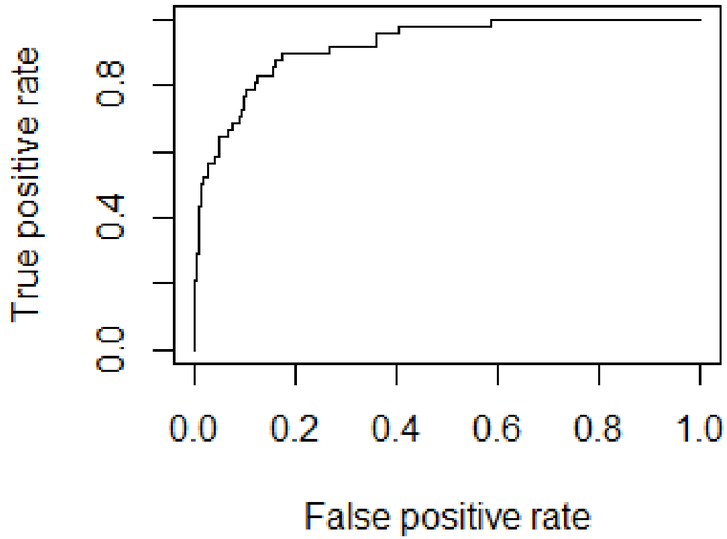
18 ×

Questão 2

Rain in Autralia



Regressão Logística
AUC 0,9233



Questões?