

SIN5007

Setembro/2017

Marcelo Figueiredo de Almeida
Tiago de Souza Oliveira

Dataset – Bank Marketing

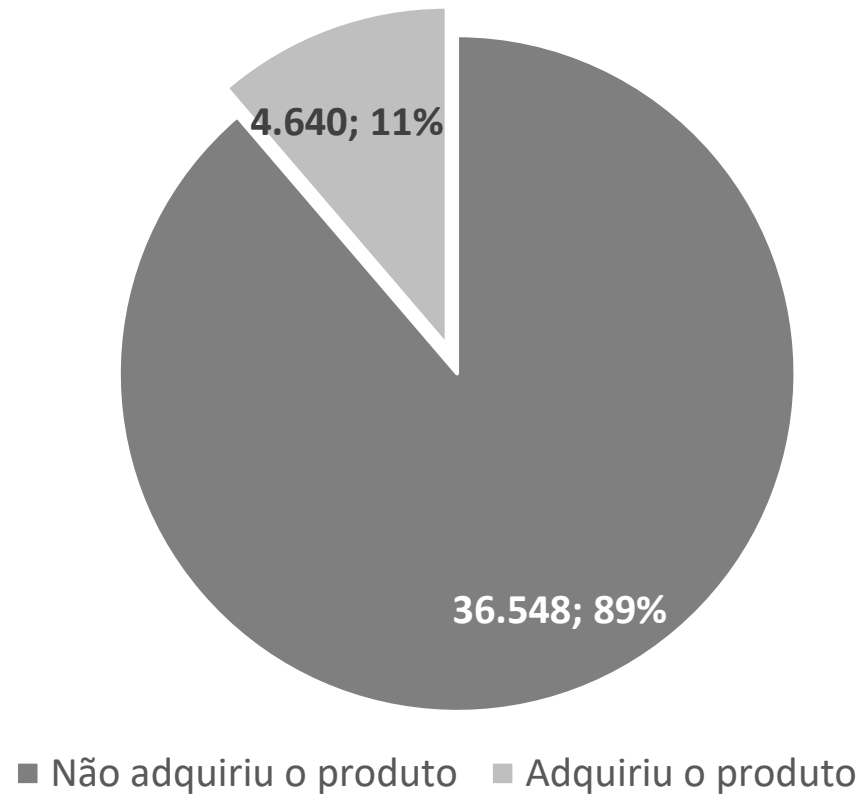
- Dados de uma instituição bancária Portuguesa;
- Campanha de marketing baseada em chamadas telefônicas;
- Produto de investimento (espécie de CDB);
- Mede a efetividade da campanha.

Dataset – Bank Marketing

- Número de observações: 41.188
- Variáveis explicativas: 16
- Variável resposta: Cliente adquiriu o produto / Não adquiriu o produto

Dataset – Bank Marketing

- Variável resposta



Dataset – Bank Marketing

Variável	Escala	Descritivo	Missings
Idade	Racional	Min 17 / Max 98	-
Trabalho	Nominal	11 categorias	330
Estado civil	Nominal	3 categorias	80
Escolaridade	Nominal	7 categorias	1.731
Inadimplente (atual)	Nominal	2 categorias	8.597
Financiamento imobiliário (atual)	Nominal	2 categorias	990
Empréstimo pessoal (atual)	Nominal	2 categorias	990
Tipo de contato	Nominal	2 categorias	-

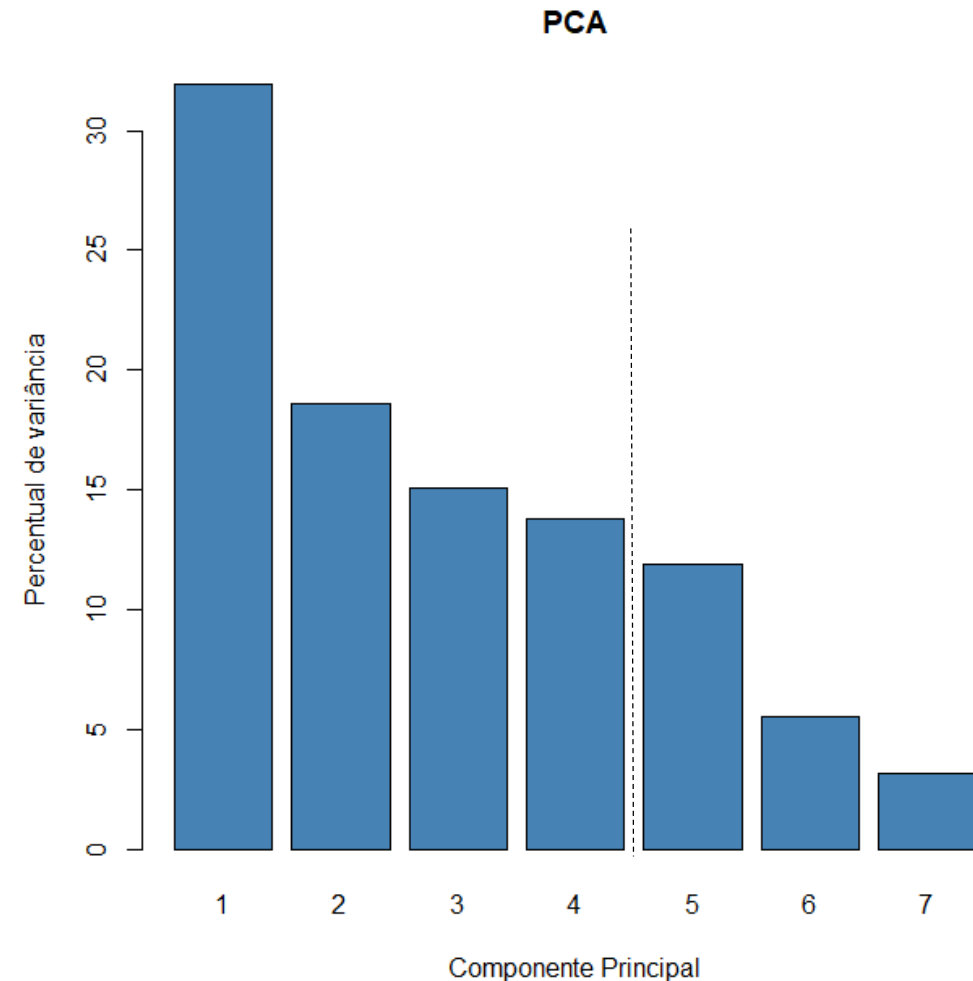
Dataset – Bank Marketing

Variável	Escala	Descritivo	Missings
Número de contatos anteriores	Racional	Min 0 / Max 55	-
Dias desde o último contato (campanha anterior)	Racional	Min 0 / Max 27	-
Número de contatos antes da campanha	Racional	Min 0 / Max 7	-
Resultado da campanha anterior	Nominal	3 categorias	-
Taxa de variação de emprego (macro)	Racional	Min -3, 4 / Max 1, 4	-
Índice de preços ao consumidor (macro)	Racional	Min 92, 2 / Max 94, 77	-
Índice de confiança do consumidor (macro)	Racional	Min -50, 8 / Max -26, 9	-
Taxa de juros (macro)	Racional	Min 0, 63 / Max 5, 04	-

Resultados - Análise de componentes principais

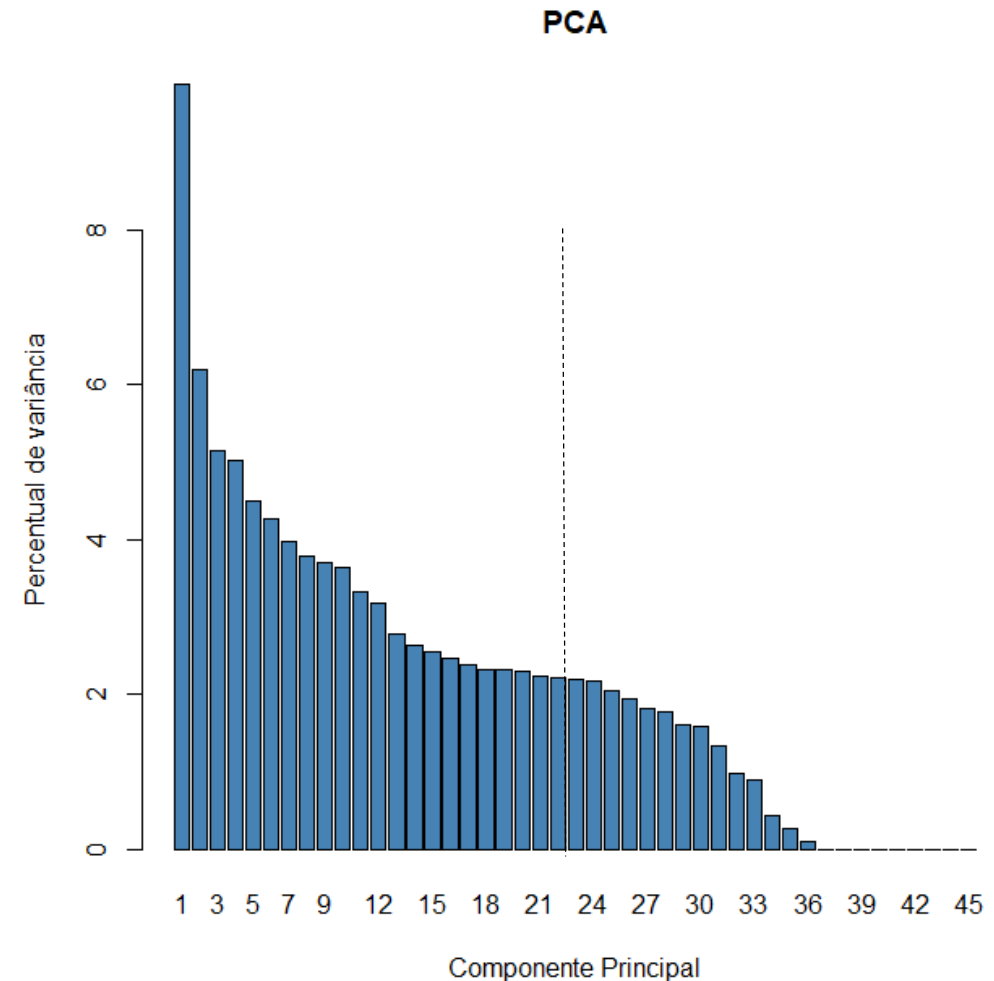
- Análise realizada nas 7 variáveis contínuas do dataset.
- Os 4 primeiros componentes explicam aproximadamente 80% da variabilidade das variáveis.

	Autovalores	Percentual de variância	Percentual de variância acumulada
comp 1	2,23	31,9%	31,9%
comp 2	1,30	18,6%	50,5%
comp 3	1,06	15,1%	65,6%
comp 4	0,96	13,8%	79,4%
comp 5	0,84	11,9%	91,3%
comp 6	0,39	5,5%	96,8%
comp 7	0,22	3,2%	100,0%



Resultados - Análise de componentes principais

- Análise realizada com todas as variáveis explicativas do dataset.
- As variáveis categóricas foram transformadas em “dummies”, dessa maneira totalizando 45 variáveis,
- O primeiro componente representa aproximadamente 9,88% da variabilidade dos dados.
- Os 22 primeiros componentes explicam aproximadamente 80% da variabilidade das variáveis.



Seleção de características - Relief

Para realizar a seleção de variáveis, aplicamos as funções *Relief* e *cutoff.k* da biblioteca *Fselector* do R.

Optamos em selecionar as variáveis que ficaram com os pesos positivos no Relief.

Selecionamos 12 variáveis, aproximadamente 66% das variáveis explicativas.

Ranking	Variável
1	educacao
2	indice_precos_macro
3	indice_confianca_macro
4	emprestimo_pessoal
5	taxa_juros_macro
6	variacao_emprego_macro
7	idade
8	num_contatos_campanha
9	tipo_contato
10	num_dias_contato_campanha_anterior
11	num_contatos_antes_campanha
12	resultado_campanha_anterior

SVM

Característica	Kernel	Gamma	Coef0	Degree	Custo	Erro	Sensibilidade	Precisão
Todas	Polinomial	1	0	3	1	0,100	0,990	0,906
PCA	Polinomial	2	1	3	1	0,102	0,991	0,904
Todas	Radial	1	-	-	1	0,102	0,985	0,908
PCA	Polinomial	1	1	3	3	0,102	0,992	0,902
Relief	Polinomial	1	1	3	1	0,103	0,973	0,910
PCA	Polinomial	1	1	3	2	0,103	0,992	0,902
PCA	Polinomial	1	1	3	1	0,103	0,993	0,901
Todas	Radial	1	-	-	3	0,104	0,982	0,908
PCA	Polinomial	2	1	3	2	0,104	0,993	0,900
PCA	Polinomial	1	0	3	1	0,106	0,996	0,896

Naive Bayes

Característica	Erro	Sensibilidade	Precisão
Todas	0.120	0.952	0.916
PCA	0.110	0.971	0.910
Relief	0.120	0.951	0.916

Rede Neural

Característica	Neurônios	Taxa de aprendizado	Erro	Sensibilidade	Precisão
Total	6	0.6	0,099	0,984	0,911
Relief	6	0.6	0,100	0,986	0,909
Total	12	0.6	0,100	0,984	0,911
Relief	8	0.6	0,100	0,984	0,910
Relief	12	0.6	0,100	0,983	0,911
Total	6	0.3	0,101	0,983	0,911
Total	4	0.1	0,101	0,982	0,911
Relief	6	0.3	0,101	0,984	0,910
Relief	12	0.3	0,101	0,983	0,911
Total	8	0.6	0,101	0,984	0,910

Melhores classificadores

Erro – Rede neural

Característica	Neurônios	Taxa de aprendizado	Erro	Sensibilidade	Precisão
Total	6	0.6	0,099	0,984	0,911

Sensibilidade – SVM

Característica	Kernel	Gamma	Coef0	Degree	Custo	Erro	Sensibilidade	Precisão
PCA	Polinomial	1	0	3	1	0,106	0,996	0,896

Precisão – Naive Bayes

Característica	Erro	Sensibilidade	Precisão
Todas	0.120	0.952	0.916