



Universidade do Minho

Bank Marketing

Aprendizagem Automática II

Bruno Veloso, A78352 | Carolina Cunha, A80142 | João Gomes, A82238



Universidade do Minho

Descrição do conjunto de dados

Cliente

- **age** – Idade do cliente (**variável numérica**);
- **job** – Tipo de emprego (**variável categórica**: admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown);
- **marital** – Estado conjugal (**variável categórica**: divorced, married, single, unknown);
- **education** – Nível de Educação (**variável categórica**: basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown);
- **default** – Crédito em Incumprimento (**variável categórica**: yes, no, unknown);
- **housing** – Empréstimo de habitação (**variável categórica**: yes, no, unknown);
- **loan** - Empréstimo pessoal (**variável categórica**: yes, no, unknown);

Descrição do conjunto de dados

Último contacto com a campanha atual

- **contact** – Tipo de comunicação (**variável categórica**);
- **month** – Último mês onde se realizou contacto (**variável categórica**);
- **day_of_week** – Último dia da semana em que se realizou contacto (**variável categórica**);
- **duration** – Duração do último contacto, em segundos (**variável numérica**);

Atributos de contextos sociais e económicos

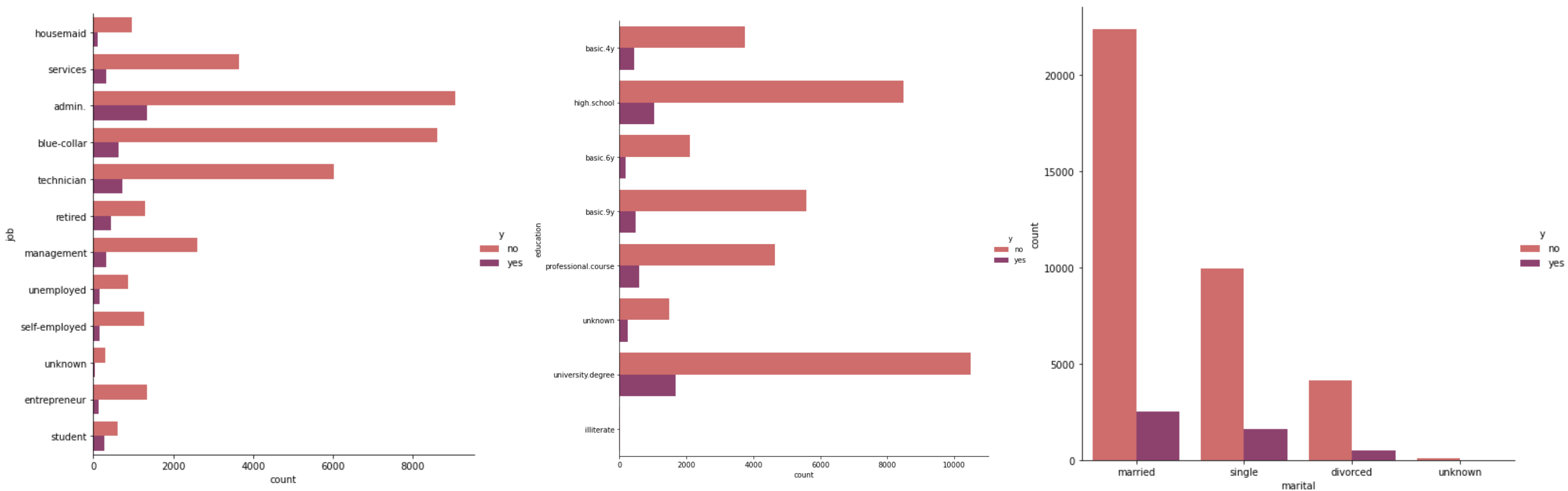
- **emp.var.rate** – Taxa de variação de emprego, indicador trimestral (**variável numérica**);
- **cons.price.idx** – Índice de preço no consumidor, indicador mensal (**variável numérica**);
- **cons.conf.idx** – Índice de confiança no consumidor, indicador mensal (**variável numérica**);
- **euribor3m** – Taxa da euribor a três meses, indicador diário (**variável numérica**);
- **nr.employed** – Número de empregados, indicador trimestral (**variável numérica**);

Outros Atributos

- **campaign** – Número de contactos realizados com um determinado cliente durante a campanha atual (**variável numérica**);
- **pdays** – Número de dias desde o último contacto com a campanha anterior (**variável numérica**, em que 999 indica que o cliente não foi contactado anteriormente);
- **previous** – Número de vezes que um cliente foi contactado antes da atual campanha (**variável numérica**);
- **poutcome** – Resultado da campanha de marketing anterior (**variável categórica**);

Análise exploratória dos dados

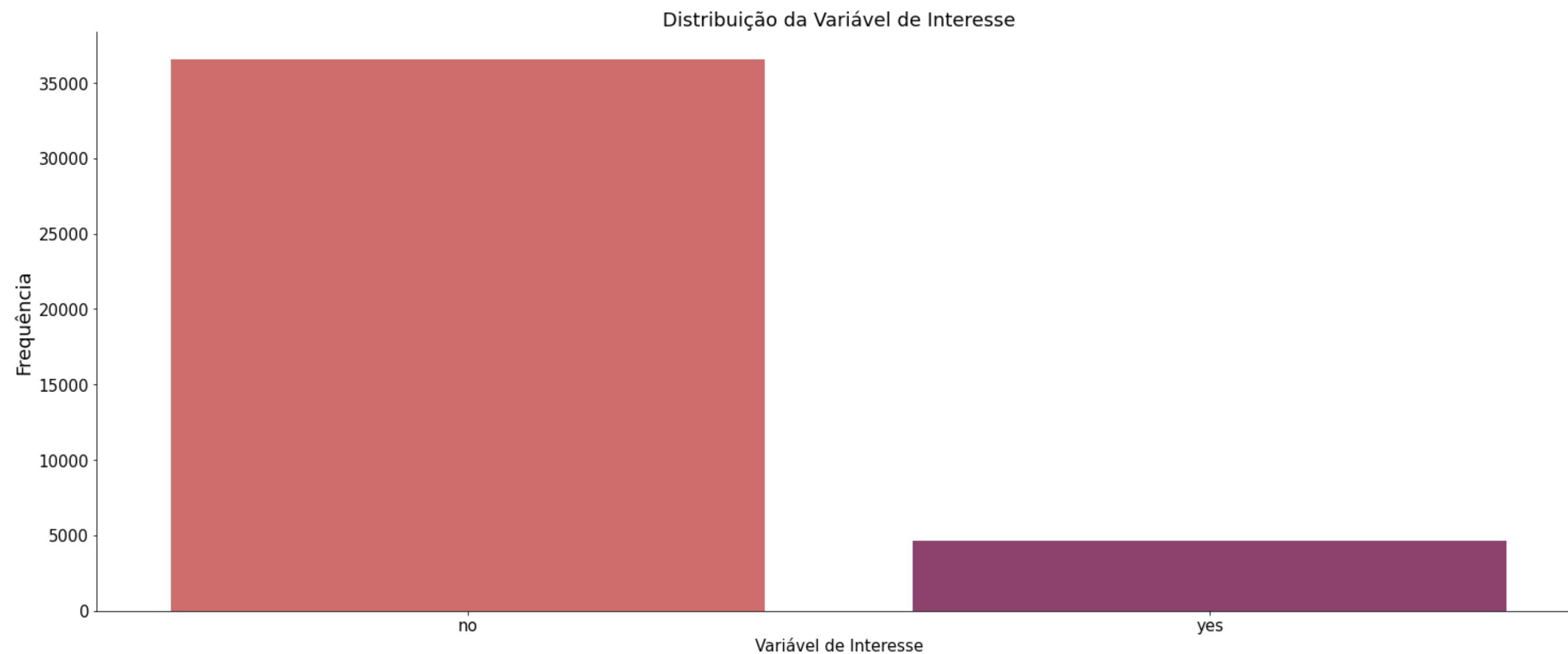
Correlação entre as variáveis

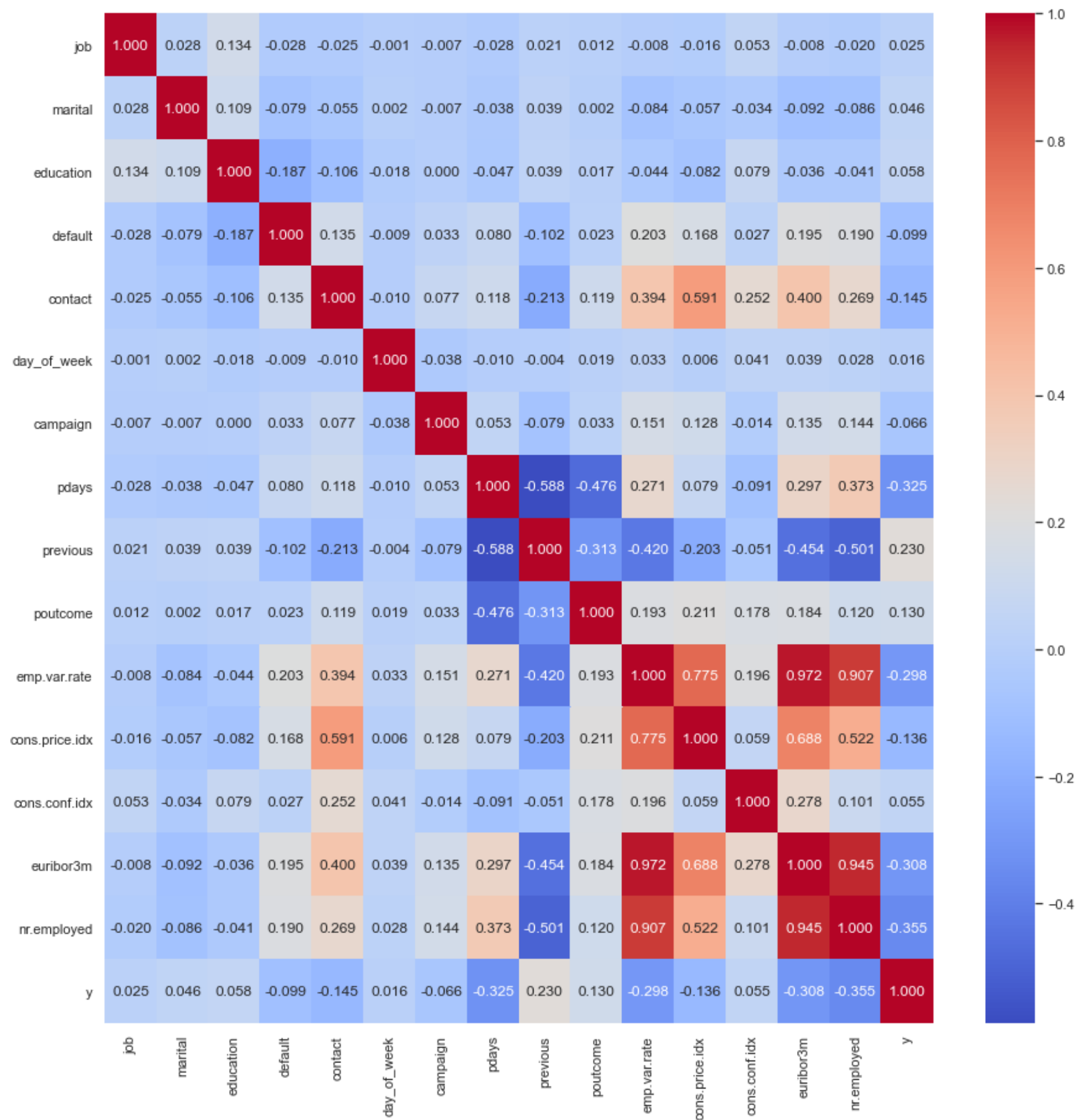




Universidade do Minho

Distribuição da variável de interesse





Matriz de correlação entre os valores numéricos



Universidade do Minho

Pré- processamento dos dados

IDADES AGRUPADAS POR FAIXA ETÁRIA

1	2	3	4
≤ 19	$>19 \ \&\& \ \leq 30$	$>30 \ \&\& \ \leq 60$	$>60 \ \&\& \ \leq 98$

DURAÇÃO AGRUPADA POR INTERVALOS

1	2	3	4
≤ 102	$>102 \ \&\& \ \leq 180$	$>180 \ \&\& \ \leq 319$	$>319 \ \&\& \ \leq 644.5$

5
>644.5



Universidade do Minho

Divisão em datasets

1

UTILIZADA A TOTALIDADE DOS DADOS

Aplicada a normalização dos dados com *StandardScaler*

3

UTILIZADA A TOTALIDADE DOS DADOS EXCETUANDO VALORES NULOS

Aplicada a normalização dos dados com *SMOTE*

5

KNN_IMPUTER NOS VALORES NULOS

Aplicada a normalização dos dados com *StandardScaler*

2

UTILIZADA A TOTALIDADE DOS DADOS EXCETUANDO VALORES NULOS

Aplicada a normalização dos dados com *StandardScaler*

4

REMOVIDAS AS COLUNAS COM POUCA CORRELAÇÃO E VALORES NULOS

age, loan, month, housing, duration. Aplicada a normalização dos dados com *StandardScaler*

Resultados Obtidos

Decision Tree (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.92	0.71	0.98	0.37	0.95	0.49	0.912
M2	0.93	0.61	0.95	0.53	0.94	0.57	0.898
M3	0.96	0.59	0.93	0.71	0.94	0.64	0.898
M4	0.91	0.63	0.98	0.24	0.95	0.35	0.900
M5	0.94	0.63	0.95	0.55	0.94	0.59	0.901

Resultados Obtidos

Gradient Boosting (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.94	0.66	0.97	0.50	0.95	0.57	0.915
M2	0.93	0.68	0.96	0.52	0.95	0.59	0.909
M3	0.95	0.61	0.94	0.68	0.94	0.64	0.902
M4	0.91	0.64	0.98	0.26	0.95	0.37	0.902
M5	0.94	0.65	0.95	0.57	0.95	0.61	0.906

Resultados Obtidos

K-Nearest-Neighbors (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.92	0.69	0.98	0.29	0.95	0.4	0.905
M2	0.90	0.71	0.98	0.26	0.94	0.38	0.893
M3	0.93	0.55	0.94	0.50	0.93	0.53	0.883
M4	0.91	0.64	0.98	0.22	0.95	0.33	0.900
M5	0.91	0.65	0.97	0.34	0.94	0.45	0.892

Resultados Obtidos

Logistic Regression (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.92	0.68	0.98	0.37	0.95	0.48	0.910
M2	0.92	0.68	0.97	0.38	0.94	0.49	0.900
M3	0.94	0.59	0.94	0.56	0.94	0.58	0.893
M4	0.91	0.69	0.99	0.19	0.95	0.30	0.901
M5	0.92	0.67	0.97	0.42	0.94	0.52	0.899

Resultados Obtidos

Naive Bayes

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.95	0.38	0.87	0.60	0.91	0.46	0.843
M2	0.99	0.16	0.24	0.97	0.39	0.27	0.336
M3	0.98	0.15	0.20	0.98	0.33	0.27	0.300
M4	0.93	0.34	0.88	0.49	0.91	0.40	0.839
M5	0.99	0.16	0.24	0.99	0.39	0.28	0.336

Resultados Obtidos

Random Forest (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.93	0.69	0.98	0.41	0.95	0.51	0.913
M2	0.92	0.68	0.97	0.45	0.95	0.54	0.904
M3	0.94	0.61	0.94	0.62	0.94	0.62	0.900
M4	0.91	0.66	0.99	0.20	0.95	0.31	0.900
M5	0.93	0.66	0.96	0.51	0.95	0.57	0.903

Resultados Obtidos

Support Vector Machines (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.93	0.66	0.97	0.40	0.95	0.50	0.909
M2	0.92	0.66	0.97	0.45	0.94	0.53	0.901
M3	0.94	0.57	0.94	0.57	0.94	0.57	0.888
M4	0.91	0.64	0.99	0.21	0.95	0.32	0.899
M5	0.91	0.68	0.98	0.36	0.94	0.47	0.896

Resultados Obtidos

XGBoost (utilizando o *GridSearchCV*)

Modelos	<i>Precision 0</i>	<i>Precision 1</i>	<i>Recall 0</i>	<i>Recall 1</i>	<i>f1_score 0</i>	<i>f1_score 1</i>	<i>Accuracy</i>
M1	0.94	0.68	0.97	0.50	0.95	0.58	0.918
M2	0.93	0.66	0.96	0.53	0.95	0.59	0.906
M3	0.95	0.63	0.94	0.64	0.94	0.63	0.904
M4	0.91	0.67	0.99	0.21	0.95	0.33	0.901
M5	0.93	0.64	0.96	0.54	0.94	0.59	0.902



Universidade do Minho

Resultados obtidos

1

XGBoost

2

Gradient Boosting

3

Gradient Boosting

4

Gradient Boosting

5

Gradient Boosting



Universidade do Minho

Próximos passos

**MODELOS NÃO
SUPERVISIONADOS**

K-Means-Clustering

Birch

DEEP LEARNING



Universidade do Minho

Bank Marketing

Aprendizagem Automática II

Bruno Veloso, A78352 | Carolina Cunha, A80142 | João Gomes, A82238