

Bank Marketing

Yuri Santos

2024-12-12

Introdução

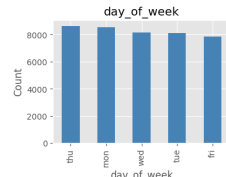
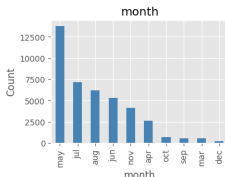
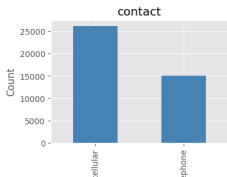
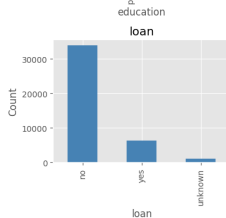
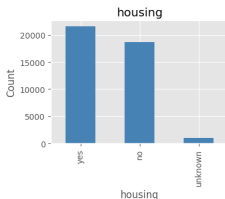
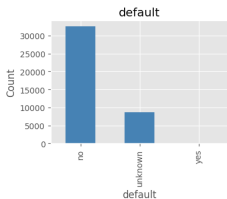
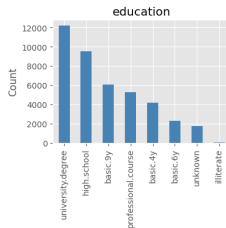
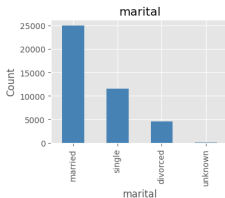
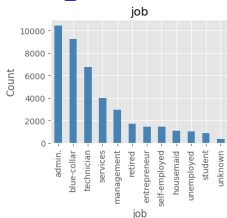
Este trabalho foca na modelagem preditiva de dados provenientes de campanhas de marketing direto realizadas por uma instituição bancária portuguesa, disponibilizados pelo UC Irvine Machine Learning Repository. Essas campanhas, baseadas em contatos telefônicos, buscavam identificar a propensão de clientes à subscrição de depósitos a prazo, um produto estratégico no setor financeiro.

Dados

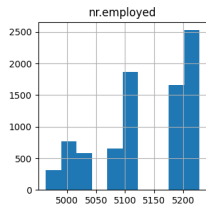
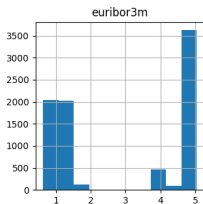
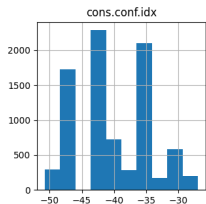
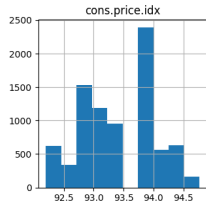
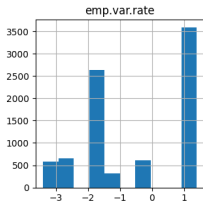
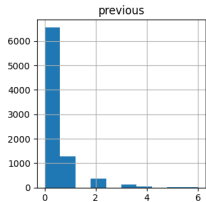
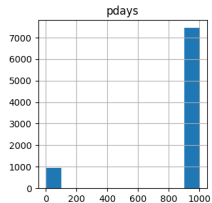
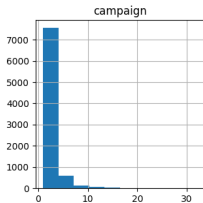
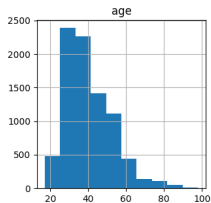
Essencialmente, há três tipos de dados.

- ▶ **Dados cadastrais**
- ▶ **Dados comportamentais**
- ▶ **Dados socio-econômicos**

Dados categóricos

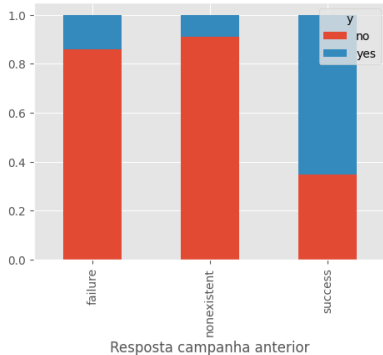


Dados numéricos



Dados resposta

(A) Relação da campanha anterior com a atual



(B) Histograma da campanha anterior

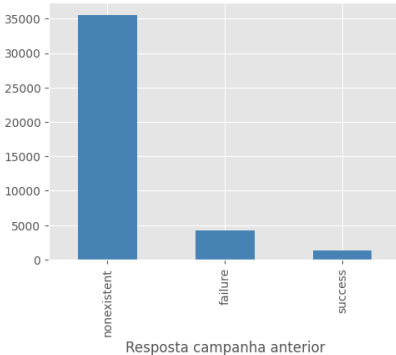


Figure 3: Distribuição e relação da variável de campanha

Pré-processamento

Devido à sub-representação das variáveis-resposta, foram criadas duas amostras de treinamento: uma com 80% dos dados, denominada *Amostra A*, e outra, de tamanho fixo (8.382 observações), balanceada com igual número de ocorrências de conversões bem-sucedidas e mal-sucedidas, denominada *Amostra B*.

A variável de inadimplência foi removida devido ao seu baixo número de ocorrências, o que dificultava a correta segmentação das amostras. Adicionalmente, a variável de duração da ligação foi excluída por ser considerada uma variável *leaky*, ou seja, que não estaria disponível no momento da predição, uma vez que seu valor é conhecido apenas após o término da interação.

As variáveis foram separadas em categóricas e numéricas para o tratamento adequado. As variáveis categóricas passaram por um processo de *one-hot encoding*, com a exclusão de pelo menos um rótulo para evitar a armadilha das variáveis *dummy*. As variáveis numéricas foram normalizadas.

Modelos e treinamento

Table 1: Modelos treinados

Modelo	Dados de treino	Estrutura	Função de Ativação	Dropout
Modelo 0	Amostra A	(49, 1)	(I, S)	(0, 0)
Modelo 1	Amostra B	(49, 40, 32, 1)	(I, S, S, S)	(0, 0, 0, 0)
Modelo 2	Amostra A	(49, 49, 49, 1)	(I, S, S, S)	(0, 0.25, 0.25, 0)
Modelo 3	Amostra B	(49, 40, 32, 1)	(I, S, S, S)	(0.2, 0.2, 0, 0)
Modelo 4	Amostra B	(49, 40, 32, 1)	(I, R, R, S)	(0.2, 0.2, 0, 0)

Modelos e treinamento

Os treinamentos foram realizados utilizando coortes de 20% dos dados para validação. O Modelo 0 e o Modelo 1 foram treinados por 25 épocas, enquanto os Modelos 2, 3 e 4 foram treinados por 75 épocas, de modo a otimizar o aprendizado e avaliar o impacto de diferentes configurações de estrutura e hiperparâmetros.

Avaliação

Table 2: Métricas de avaliação com dados de teste

Modelo	Acurácia	Perda (<i>Loss</i>)
Modelo 0	0,900	0,278
Modelo 1	0,580	0,848
Modelo 2	0,901	0,273
Modelo 3	0,695	0,768
Modelo 4	0,692	0,755

Avaliação

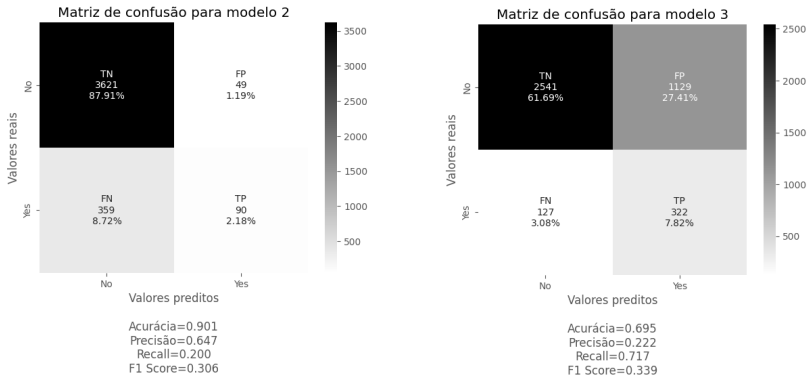


Figure 4: Matrizes de confusão do Modelo 2 e Modelo 3, acompanhadas de outras métricas

Avaliação

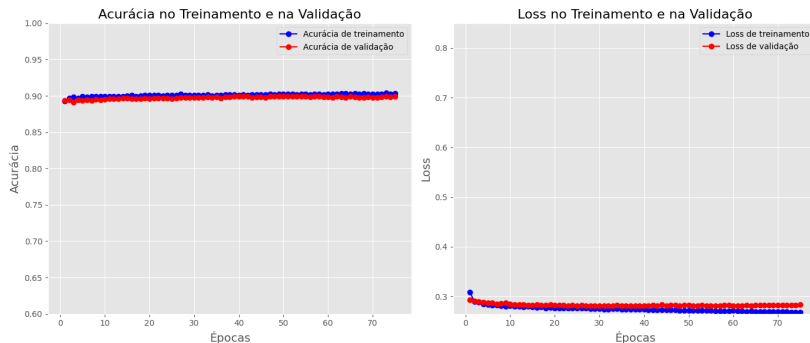


Figure 5: Acurácia e loss durante o treinamento e validação do modelo 2

Avaliação

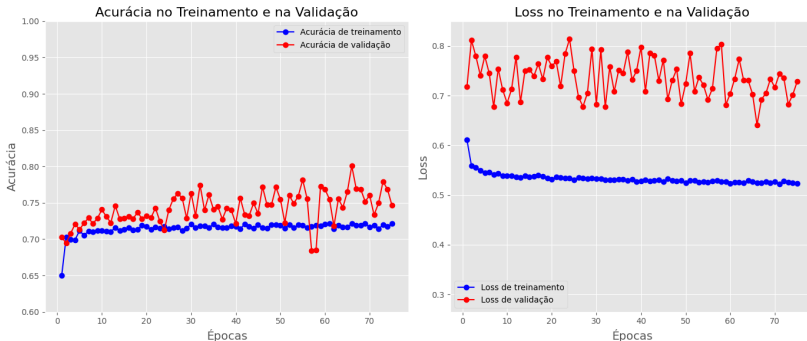


Figure 6: Acurácia e loss durante o treinamento e validação do modelo 3

Conclusão

- ▶ Escolhas de modelo a partir do contexto
- ▶ Enriquecimento de dados
- ▶ Estratégias de aprimoramento

Referências

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.

Ferreira, J. A.. Tópicos Especiais em Estatística Computacional. (2024, Dezembro). UFPE.

https://www.de.ufpe.br/~jodavid/material/topicos_especiais_est_compu

Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of the European Simulation and Modelling Conference-ESM* (Vol. 2011).

Moro, S., Rita, P., & Cortez, P. (2014). Bank Marketing [Dataset]. UCI Machine Learning Repository.
<https://doi.org/10.24432/C5K306>.

Verbraken, T., Verbeke, W., & Baesens, B. (2014). Profit optimizing customer churn prediction with Bayesian network classifiers. *Intelligent Data Analysis*, 18(1), 3-24.