



NEUROTECH

LEOMAR FONSECA

DESAFIO PRÁTICO

CIENTISTA DE DADOS II

ÍNDICE

- 1 Contexto do desafio
- 2 Abordagem e boas práticas utilizadas
- 3 Resultados do modelo
- 4 Análise financeira
- 5 Sugestões ou próximos passos

CONTEXTO

Desenvolver um modelo de concessão de crédito para uma base de dados desbalanceada. O modelo deve usar a variável TARGET como resposta, para avaliar se um ID específico deverá ser considerado bom ou mau pagador. Além disso, as variáveis estão mascaradas, com exceção da REF_DATA, IDADE e das variáveis já mencionadas.

```
1 # Não existem valores repetidos de ID
2 df_train['ID'].nunique()
```

✓ 0.5s

120750

```
1 # Variável target
2 print(f'Distribuição da variável target:')
3 df_train['TARGET'].value_counts(normalize=True)
```

✓ 0.5s

Distribuição da variável target:

TARGET

0 0.754973

1 0.245027

Name: proportion, dtype: float64



ABORDAGEM


E BOAS PRÁTICAS UTILIZADAS

- GIT PARA VERSIONAMENTO DO CÓDIGO
- DOCUMENTAÇÃO COM COMENTÁRIOS AO LONGO DO DESENVOLVIMENTO
- *VIRTUAL ENVIRONMENT*
- CUIDADO COM DATA LEAKAGE
- DEFINIÇÃO DE VARIÁVEL PARA REPRODUCIBILIDADE (SEED)



RESULTADOS

O modelo demonstrou uma boa performance no dataframe de treino, mas não conseguiu generalizar bem as características para ficar acima dos “números mágicos” do dataframe de teste.



Parâmetro observado	Dataframe de treino	Dataframe de teste	Número mágico
Recall	0.50	0.29	-
F1-score	0.58	0.33	-
AUC	0.87	0.63	0.60
Gini	0.73	0.27	0.50
KS	0.55	0.19	0.3

ANÁLISE FINANCEIRA



```
1 # Pergunta 1
2 print(f"A financeira emprestou R${1000 * df_test[df_test['ref_date_trunc'] == '2017-08-01'].shape[0]}")
```

✓ 0.1s

A financeira emprestou R\$7305000

```
1 # Pergunta 2
2 print(f"O valor da dívida total é R${1000 * df_test[df_test['predicted-as-is'] == 1][df_test['ref_date_trunc'] == '2017-08-01'].shape[0]}")
```

✓ 0.1s

O valor da dívida total é R\$1537000

```
1 # Pergunta 3.1
2 print(f"O percentual de pessoas que teve o empréstimo negado foi: {df_test[df_test['ref_date_trunc'] == '2017-08-01']['predicted-as-is'].mean()*100 :.2f}%")
```

✓ 0.1s

O percentual de pessoas que teve o empréstimo negado foi: 21.04%

```
1 # Pergunta 3.2
2 print(f"Definição do threshold para a política to-be: {df_test['probability_default'].quantile(0.2104) * 100 :.2f}%")
```

✓ 0.7s

Definição do threshold para a política to-be: 4.57%



```
1 # Política TO BE
2 df_test['predicted-to-be'] = np.where(df_test['probability_default'] > 0.0457, 1, 0)
3 print(f"O valor da dívida total na política to be seria R${1000 * df_test[df_test['predicted-to-be'] == 1][df_test['ref_date_trunc'] == '2017-08-01'].shape[0]}")
```

✓ 0.2s

O valor da dívida total na política to be seria R\$5517000



PRÓXIMOS PASSOS

REFINAMENTO E
INDUSTRIALIZAÇÃO



- BINNING DE VARIÁVEIS
- DATASET DE VALIDACAO
- HYPERPARAMETER TUNING
- CRIAR VARIÁVEIS HISTÓRICAS E DE TENDÊNCIA
USANDO DADOS EXTERNOS
- SCORAR O OOT EM OUTRO ARQUIVO COM UMA
PIPELINE SERIALIZADA DO TRATAMENTO E DO
MODELO EM OOP
- TESTAR OUTROS MODELOS (SVM, REDES
NEURAS)



Contato:

Leomar Fonseca 
leomar.fmn@gmail.com
(81) 99999 3512

