

FIAP

NBA



Feature engineering

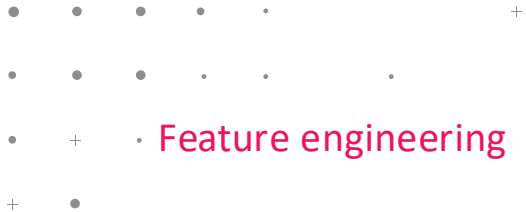
CONTEXTO E DESAFIOS

A **QuantumFinance** possui um modelo de risco de crédito desatualizado, que tem concedido cartões para um número muito alto de mau pagadores e gerando problemas financeiros para a companhia.

Neste contexto, o HEAD de Ciência de Dados chamou os melhores talentos de sua equipe para ajudá-lo na construção de um novo modelo que será utilizado antes da **concessão de cartão de crédito aos requerentes**. Uma primeira análise feita no dataset mostrou que será necessário despender grande esforço de Feature Engineering antes do desenvolvimento de fato. De qualquer maneira ele está extremamente confiante, pois sabe que possui uma excelente equipe para resolver o problema.

O desafio de sua equipe nesta etapa será realizar melhorias no dataset baseadas nas melhores práticas de Feature Engineering para que, num momento posterior, seja feito o desenvolvimento do modelo.

Boa sorte!

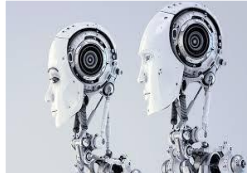


+

A melhoria do dataset deverá ser feita com base nas técnicas fundamentais de *Feature Engineering* como imputação de dados, normalização, codificação de variáveis categóricas, seleção de características, etc.

• +

-
- A 4x4 grid of 16 squares, each containing a different geometric shape or pattern. The shapes include squares, circles, triangles, and combinations thereof, some with internal lines or shading.



Feature engineering

ENTREGÁVEIS

Ao término do projeto a equipe deverá fazer as seguintes entregas:

Arquivo powerpoint com análises e resultados:

A apresentação deverá conter não apenas o resultado final obtido e uma explicação detalhada das decisões tomadas para atingir o resultado alcançado.

Em particular, alguns pontos são extremamente importantes:

- Detalhamento de todas as Técnicas de Feature Engineering aplicadas no dataset com justificativas das decisões tomadas;
- Resultados intermediário que a equipe julgar importante apresentar para demonstrar a evolução de qualidade do dataset;
- Métricas de avaliação do modelo, deixando claro o impacto das técnicas de feature

Jupyter Notebook:

O notebook deverá estar bem organizado com seções claras e textos que facilitem a compreensão da análise e decisões tomadas e que permita a obtenção do resultado final a partir do dataset disponibilizado.

Importante: parte significativa da avaliação do resultado será feita com a execução do notebook. Desta forma, é importante que todas as células executem corretamente e que os resultados sejam integralmente reprodutíveis;

Feature engineering

DATA SET - Dicionário de dados

Var. Id	Característica	Descrição
1	ID_CLIENTE	Número sequencial para o requerente (a ser usado como chave)
2	TIPO_FUNCIONARIO	Não informado
3	DIA_PAGAMENTO	Dia do mês para o pagamento da fatura, escolhido pelo requerente
4	TIPO_ENVIO_APLICACAO	Indica se a aplicação foi submetida via internet ou pessoalmente/enviada pelo correio
5	QUANT_CARTOES_ADICIONAIS	Quantidade de cartões adicionais solicitados no mesmo formulário de aplicação
6	TIPO_ENDERECO_POSTAL	Indica se o endereço para envio é o endereço residencial ou outro. Codificação não informada.
7	SEXO	
8	ESTADO_CIVIL	Codificação não informada
9	QUANT_DEPENDENTES	
10	NIVEL_EDUCACIONAL	Nível educacional em ordem gradual não informado
11	ESTADO_NASCIMENTO	
12	CIDADE_NASCIMENTO	
13	NACIONALIDADE	País de nascimento. Codificação não informada, mas é provável que o Brasil seja igual a 1.
14	ESTADO_RESIDENCIAL	Estado de residência
15	CIDADE_RESIDENCIAL	Cidade de residência
16	BAIRRO_RESIDENCIAL	Bairro de residência
17	FLAG_TELEFONE_RESIDENCIAL	Indica se o requerente possui um telefone residencial
18	CODIGO_AREA_TELEFONE_RESIDENCIAL	Pseudo-código de três dígitos
19	TIPO_RESIDENCIA	Codificação não informada. Em geral, existem tipos como: própria, hipotecada, alugada, dos pais, familiar, etc.
20	MESES_RESIDENCIA	Tempo na residência atual em meses
21	FLAG_TELEFONE_MOVEL	Indica se o requerente possui um telefone móvel
22	FLAG_EMAIL	Indica se o requerente possui um endereço de e-mail
23	RENDA_PESSOAL_MENSAL	Renda mensal regular pessoal do requerente em moeda brasileira (R\$)
24	OUTRAS_RENDAS	Outras rendas mensais médias do requerente em moeda brasileira (R\$)
25	FLAG_VISA	Indicador se o requerente é titular de um cartão de crédito VISA

Feature engineering

DATA SET - Dicionário de dados

Var. Id	Característica	Descrição
26	FLAG_MASTERCARD	Indicador se o requerente é titular de um cartão de crédito MASTERCARD
27	FLAG_DINERS	Indicador se o requerente é titular de um cartão de crédito DINERS
28	FLAG_AMERICAN_EXPRESS	Indicador se o requerente é titular de um cartão de crédito AMERICAN EXPRESS
29	FLAG_OUTROS_CARTOES	Apesar de ser rotulado como "FLAG", esse campo apresenta três valores não explicados
30	QUANT_CONTAS_BANCARIAS	
31	QUANT_CONTAS_BANCARIAS_ESPECIAIS	
32	VALOR_PATRIMONIO_PESSOAL	Valor total dos bens pessoais, como casas, carros, etc., em moeda brasileira (R\$)
33	QUANT_CARROS	Quantidade de carros que o requerente possui
34	EMPRESA	Se o requerente forneceu o nome da empresa onde ele/ela trabalha formalmente
35	ESTADO_PROFISSIONAL	Estado onde o requerente trabalha
36	CIDADE_PROFISSIONAL	Cidade onde o requerente trabalha
37	BAIRRO_PROFISSIONAL	Bairro onde o requerente trabalha
38	FLAG_TELEFONE_PROFISSIONAL	Indica se o número de telefone profissional foi fornecido
39	CODIGO_AREA_TELEFONE_PROFISSIONAL	Pseudo-código de três dígitos
40	MESES_NO_TRABALHO	Tempo no emprego atual em meses
41	CODIGO_PROFISSAO	Código profissional do requerente. Codificação não informada
42	TIPO_OCUPACAO	Codificação não informada
43	CODIGO_PROFISSAO_CONJUGE	Código profissional do cônjuge. Codificação não informada
44	NIVEL_EDUCACIONAL_CONJUGE	Nível educacional do cônjuge em ordem gradual não informado
45	FLAG_DOCUMENTO_RESIDENCIAL	Indicador de confirmação documental do endereço residencial
46	FLAG_RG	Indicador de confirmação documental do número do cartão de cidadão
47	FLAG_CPF	Indicador de confirmação documental do status de contribuinte
48	FLAG_COMPROVANTE_RENDA	Indicador de confirmação documental da renda
49	PRODUTO	Tipo de produto de crédito aplicado. Codificação não informada
50	FLAG_REGISTRO_ACSP	Indicador se o requerente possui alguma inadimplência de crédito anterior
51	IDADE	Idade do requerente no momento da inscrição
52	CEP_RESIDENCIAL_3	Três dígitos mais significativos do CEP residencial atual
53	CEP_PROFISSIONAL_3	Três dígitos mais significativos do CEP do emprego atual
54	ROTULO_ALVO_MAU=1	Variável Alvo: BAD=1, BOM=0



DATASET

REF.: <https://raw.githubusercontent.com/diogenesjusto/FIAP/master/dados/credit.csv>

```
df = pd.read_csv("https://raw.githubusercontent.com/diogenesjusto/FIAP/master/dados/credit.csv", encoding='unicode_escape',  
sep='\t', header=None)
```

FIAP