

Desafio de Inteligência Artificial Blu

Seja muito bem-vindo(a) ao teste para fazer parte do nosso time. Esse teste é composto por algumas questões para serem respondidas/desenvolvidas. Queremos conhecer melhor suas habilidades matemáticas e de tecnologia antes de fazer parte do nosso time.

Podemos dizer que passamos boa parte do tempo programando em Python e fazendo queries, então se você puder mostrar o que sabe sobre esses assuntos, nos ajudará muito. Não precisa responder todas as questões, nem fazer do jeito sugerido, mas quanto mais próximo do nosso esperado melhor. Nos surpreenda.

Sugerimos que você faça um fork desse repositório no GitHub para a sua conta pessoal e, se possível, compartilhe o acesso ao repositório conosco. Gostaríamos que você abrisse uma nova branch a partir da main com o seu nome. Conforme for evoluindo seu código você pode ir fazendo commits na sua branch e, no momento que achar adequado, pode abrir uma “Pull Request” da sua branch para a main. Se você não dominar o GitHub, pode nos enviar os arquivos por email em uma pasta zipada com o conteúdo desenvolvido.

Em boa parte dos projetos utilizamos o notebook (arquivo .ipynb) pois facilita algumas análises e acelera o desenvolvimento. Sendo assim, é desejável que ao longo do teste você crie ao menos um módulo em Python e consiga consumi-lo em um notebook (arquivo .ipynb). Dentro do repositório você pode salvar todo o seu desenvolvimento, módulos e até mesmo respostas em algum outro arquivo, se desejar.

Questão 1

No arquivo .csv em anexo (ia_vende_normalmente.csv) temos dados fictícios de vendas de um determinado cliente. Gostaríamos de ver a sua linha de raciocínio e os códigos feitos para responder:

- Estime a probabilidade deste cliente ficar 29 dias consecutivos sem vender. Justifique o procedimento escolhido para realizar a estimativa.
- A partir de quantos dias consecutivos sem vendas seria plausível afirmar que o cliente deu churn? Justifique sua resposta.
- É possível perceber uma tendência nas vendas? Justifique.
- Faça uma previsão para o volume de vendas dos próximos 15 dias do cliente. Aqui, quanto maior a expertise demonstrada em temas como: escolha do modelo, pré-processamento dos dados, estimativa de erro cometido, etc., melhor.
- Supondo que vamos colocar os dados de valores das vendas na entrada de um algoritmo de predição combinado com outros dados. Como você prepararia esses dados para a entrada no algoritmo e como eles ficariam?

Questão 2

No arquivo relacionamento_ia.csv temos clientes listados que aparecem como cobradores ou cobrados.

- Como você definiria a distância de relacionamento entre clientes com estas informações?
- Com a sua definição, qual a distância entre todos os clientes?

Questão 3

Dadas as tabelas a seguir que tratam do cadastro dos clientes e de suas vendas. A tabela de vendas trás os identificadores dos clientes, seu nome e cpf_cnpj. Cada venda está relacionada a um cliente e tem um valor de venda (Gross_value) e a data em que ocorreu (created_at). Considere que nessa tabela existem mais de 1000 redes que chegam a fazer mais de uma venda por dia e temos registros de dados desde 2017.

Clients	Sales
Id	Id
Name	Gross_value
Cpf_cnpj	Client_uuid
Parent_client_id	Created_at
uuid	

Sabendo que o parent_client_id é o identificador do dono da rede. Gostaríamos de um procedimento usando SQL que retorne como resposta uma tabela com as seguintes colunas:

- nome_da_rede: nome da rede analisada como sendo a conta dona da rede
- cpf_cnpj_rede: cpf ou cnpj da rede analisada
- ano: ano de referência em que os valores foram computados para a rede
- mes: mês de referência em que os valores foram computados para a rede
- vendas_mes: valor total de vendas do mês referenciado acima
- vendas_acumulado_6_meses: valor total de vendas acumuladas dos últimos 6 meses antes do mês de referência. Incluir o mês de referência mais os 5 meses para trás.
- intervalo_medio_vendas: intervalo médio entre vendas no mês calculado em dias

Obs. 1: Os períodos a serem analisados devem ser de uma data maior ou igual a 01/01/2020.

Obs. 2: O retorno deve contemplar todos os clientes e suas respectivas datas em que houve venda.

Questão 4

Programação pura. Uma aproximação da função $f(x) = e^x$ dependendo de um número N é dada por:

$$\sum_{k=0}^N \frac{x^k}{k!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^N}{N!},$$

em que $k! = k*(k-1)*(k-2)*\dots*1$ é o fatorial do número k .

- Construa um algoritmo em Python (ou linguagem similar) que, dados x e N , retorne a aproximação descrita acima.
- Calcule o erro da aproximação de e^4 para os valores de $N = 5, 15, 20$ e 30 . Para avaliar o valor de e^4 , pode usar uma biblioteca externa.