Nome completo: **Johnny Hideki Horita**

**DESAFIO EMPIRICUS – CURSO DATA SCIENCE**

Queremos recomendar assinaturas e cursos na aba lateral do site da Empiricus para nossos já usuários.

Temos as seguintes informações dos usuários (fictícias):

* CPF
* Data de Nascimento
* CEP
* Assinaturas e Cursos ativos
* Assinaturas e Cursos cancelados
* Formas de Pagamento usadas nas compras anteriores

E temos as seguintes informações sobre as assinaturas e cursos (fictícias):

* Preço
* Categoria (Renda fixa, renda variável, internacional, crypto)
* Autor das publicações

**Baseado na nossa intenção e nos dados fictícios que já temos, redija sobre:**

1. Que tipo de dados estamos lidando?
2. Qual seria um algoritmo apropriado?
3. Qual passos você tomaria para resolver?
4. Quais dificuldades você antecipa?
5. Quais dados adicionais você acha que poderia usar?
6. Que tipo de dados estamos lidando?

Teoricamente estamos lidando com dados categóricos e quantitativos, podendo ser estruturados e ou semiestruturados.

1. Qual seria um algoritmo apropriado?

O algoritmo apropriado seria o de **recomendação**, baseado nos critérios de Assinatura e Cursos Ativos, Assinatura e Cursos cancelados, na Forma de Pagamento usadas nas compras anteriores, no preço, na categoria, no autor da publicações e outras váriáveis e com **a ajuda de uma predição probabilística** que possam definir as recomendações de novas assinaturas e cursos para os usuários.

1. Qual passos você tomaria para resolver?

Para resolver o modelo de recomendação eu aplicaria 3 análises em conjunto para definir a melhor recomendação, seriam elas:

1. Algoritmo de regras de Associação
2. Filtragem de Conteúdo
3. Filtragem Colaborativa
4. Algoritmo de regras de Associação

Com cálculos estatísticos de frequência no conjunto de dados e com auxílio de variáveis de confiança definimos algumas regras para atender um critério, com isso conseguimos realizar algumas recomendações.

Exemplo:

Analisando o conjunto de dados (dataset) com relação aos itens adiquidos pelos usuários, assinaturas e cursos, percebemos que os usuários que assinam ou compram o produto X assinam ou compram o produto Y, então a um determinado percentual de certeza (definição do critério de regra), podemos afirmar que o próximo usuário que assinar ou comprar o produto X provavelmente vai assinar ou comprar o produto Y.

1. Filtragem de Conteúdo

Na filtragem de conteúdo agrupamos as assinaturas e cursos com base em características semelhantes, como categoria, preço e o autor.

Associando a análise de Algoritmo de regras de Associação, onde percebo que os usuários que adquiriram a assinatura ou curso X também possuem a assinatura ou curso Y, dessa forma tenho uma probabilidade alta que o usuário irá gostar da recomendação da assinatura ou curso Y.

Exemplo:

Quem gostou da assinatura X da categoria A provavelmente deve gostar da assinatura Y da categoria A.

1. Filtragem Colaborativa

Consiste em analisar as semelhanças entre os usuários e, com isso, determinar uma recomendação de acordo com os usuários mais parecidos.

Identificamos que um usuário A é parecido com o usuário B, se o usuário A gostou da assinatura ou curso X, então podemos supor que B poderá gostar da assinatura ou curso X também.

Para definirmos um critério de semelhança utiizaremos as caracteristicas e o histórico de usuários, com essas informações utilizaremos modelos para calcular a relação de aproximidade entre os usuários.

Exemplo:

Quando um novo usuário C chegar para receber recomendações, o algoritmo efetuará a análise “Com base nas informações de dados do usuário C, o usuário que mais se aproxima das informações é o usuário A, então o algoritmo recomenda para C o que A gostou.”

1. Quais dificuldades você antecipa?

Abaixo listo as principais dificuldades na implemetanção dos algoritmos e as possíveis alternativas para contorná-las.

1. Conhecendo o usuário

Uma das principais dificuldades de recomendação ou experiência personalizada seria indicar as primeiras recomendações visto que não conhecemos o usuário no seu primeiro contato com a Site da Empiricus.

Nesse caso uma abordagem seria recomendar as assinaturas ou cursos, que foram mais vendidos, os que estão mais em alta.

1. Filtragem de Conteúdo
2. Falta de informação:

Para a filtragem colaborativa é necessário conhecer bem os usuários da plataforma, pois precisamos de informações qualitativas sobre as assinaturas e cursos que o usuário gostou, precisamos saber a opinião dos usuários sobre as assinaturas e cursos, seja ela positiva ou negativa.

1. Site desatualizado

Um ponto problemático que eu identifiquei foi no conteúdo dos cursos oferecidos, quando eu clico em “Conhecer curso” sou direcionado para a página principal da Empiricus, ou sou direcionado para uma página de “Lista de espera”, ou sou direcionado para uma página de Oferta Encerrada aguardo 6 segundos e sou direcionado para uma página de Nossos Produtos, mas eu estava avaliando Cursos (não faz sentido).

Este é um ponto problemático caso efetue recomendações de Cursos que não tem o seu cadastro atualizado e alinhado com a Plataforma WEB.

1. Quais dados adicionais você acha que poderia usar?