**💡 Nova Ideia de Projeto: Segmentação Inteligente de Clientes**

**🎯 Título Provisório**

**"Redefinindo a Segmentação de Clientes no Agronegócio com Clusterização: Uma Abordagem Não Supervisionada para Eliminar Categorias Indefinidas"**

**✅ Objetivo Geral**

Criar uma nova segmentação de clientes utilizando **técnicas de clusterização**, com base em dados operacionais, comerciais e comportamentais, substituindo o rótulo **IND (indefinido)** por categorias bem definidas, com valor analítico e estratégico.

**🎯 Objetivos Específicos**

* Analisar as limitações da segmentação atual, com foco no rótulo IND.
* Agrupar os clientes em segmentos **comportamentais** ou **operacionais** coerentes, com base em dados históricos.
* Gerar perfis claros para cada cluster, auxiliando áreas comercial, técnica e marketing.
* Propor uma nova forma de segmentar **todos os clientes**, inclusive os que hoje não têm rótulo definido.

**🧩 Fontes de Dados (Features para Clusterização)**

| **Fonte** | **Variáveis relevantes** |
| --- | --- |
| fHistoricoVendas | total comprado (ano/safra), ticket médio, frequência de compras, variedade de produtos |
| fEntregaGrãos | volume total entregue, % sobre potencial |
| fAstec | número de atendimentos, frequência, tempo médio entre atendimentos |
| fPlanoSafra | área plantada, tipo de cultura, potencial de produção |
| dCarteira | status atual (ativo/inativo), filial associada |
| dCadastroPropriedadesCRM/ERP | localização geográfica, tamanho da propriedade |

**🧠 Técnicas e Algoritmos**

* **Pré-processamento:** normalização, tratamento de nulos, PCA (se necessário)
* **Algoritmos de Clusterização:**
  + KMeans (bom para início)
  + DBSCAN (detecta outliers e agrupamentos não esféricos)
  + Hierárquico (Ward linkage para dendrogramas e cortes otimizados)
* **Validação dos Clusters:**
  + Silhouette Score
  + Distância intra e inter cluster
  + Visualização com PCA ou t-SNE

**📌 Saídas Esperadas**

* Um modelo de clusterização aplicável a toda base de clientes
* Substituição do rótulo IND por um novo grupo segmentado via dados
* Perfis de cliente por grupo (ex: Cluster 1 = cliente de alto potencial e alta entrega, Cluster 2 = pequeno produtor de ciclo curto, etc.)
* Dashboard com clusters e características

**🧪 Exemplos de Perfis Esperados**

| **Cluster** | **Perfil Comportamental** |
| --- | --- |
| 1 | Alto potencial de plantio, alta compra, entrega elevada |
| 2 | Baixo volume de compras, poucas entregas, sem assistência técnica |
| 3 | Médio potencial, muitos atendimentos técnicos, fidelizado |
| 4 | Cliente novo, potencial médio, ainda sem histórico expressivo |

**🧰 Ferramentas Sugeridas**

* Python (pandas, sklearn, seaborn, matplotlib, plotly)
* Power BI ou Streamlit (para exibição dos clusters)
* Jupyter Notebook (desenvolvimento do modelo)

**🔍 Questão Norteadora**

"Podemos criar uma segmentação objetiva e orientada por dados para todos os clientes, eliminando o uso do rótulo 'Indefinido', e agrupando clientes de acordo com seu comportamento de compra, entrega e relacionamento?"

**⚙️ Etapas Sugeridas do Projeto**

1. **Exploração da base existente e avaliação da coluna de segmentação**
2. **Construção do dataset de entrada (features relevantes)**
3. **Análise exploratória dos dados (EDA)**
4. **Clusterização com KMeans e comparação com outros métodos**
5. **Interpretação dos clusters e substituição dos 'IND'**
6. **Criação de documentação ou dashboard com os novos grupos**
7. **Sugestões de uso prático da nova segmentação no CRM/ERP**

Se quiser, posso te ajudar a:

* Escolher os melhores atributos para clusterizar
* Criar o código inicial com KMeans
* Fazer um exemplo de visualização de clusters
* Esboçar os perfis dos clusters com base nos dados

Quer seguir com isso? Posso começar com um esboço do dataset que você poderia montar para alimentar o modelo.