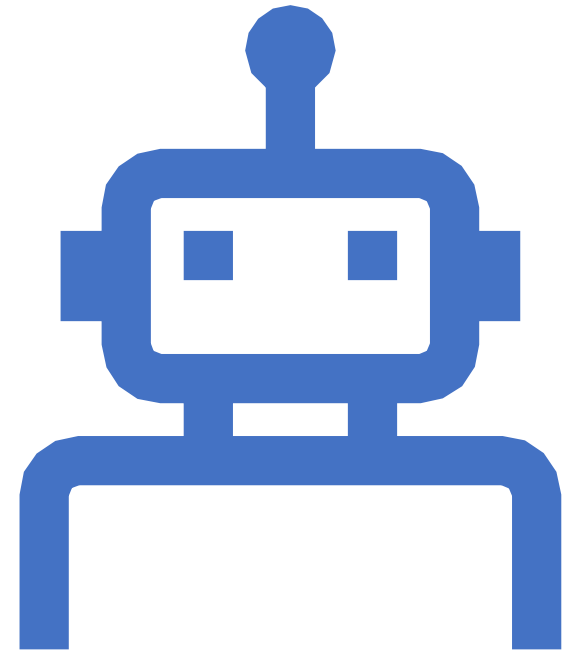


Classe "Desbalanceada"

- A maioria das transações não são fraudulentas
- A maioria dos pacientes não estão doentes
- A maioria dos clientes não vai abandonar a empresa
- A maioria dos acessos a rede não são ataques



Soluções

1. Coletar dados de forma balanceada
2. Havendo dados em abundância, gerar amostras estratificadas e balanceadas
3. Usar métricas de avaliação apropriadas
4. Utilizar modelos que penalizam o classificador ao errar a classe minoritária
5. Gerar dados artificialmente, balanceando a classe

Gerar ou Remover

- Undersampling:
 - Excluir dados da classe com maior presença
- Oversampling
 - Produzir artificialmente dados para a classe com menor presença

Gerar Dados Artificialmente

- Método Smote
 - Gera dados artificialmente utilizando o "vizinho mais próximo" da classe