

Explicações Contrastivas em Classificadores Baseados em Árvores

On Contrastive Explanations **for Tree-Based Classifiers**

Equipe: Débora de Lima Silva, Guilherme Oliveira Fernandes, Kaique Hilquias Nunes de Souza, Pedro Henrique Araújo de Oliveira



Motivação

- *Explicabilidade é crucial para IA confiável*
- *Modelos baseados em árvores (Random Forests, Boosted Trees) são precisos, mas opacos*
- *Usuários querem saber:*
 - *Por que uma previsão foi feita? (explicação abductiva)*
 - *Por que não outra classe? (explicação contrastiva)*
- *Explicabilidade é chave em áreas críticas: saúde, crédito, justiça*

Tipos de Explicações

- **Abdutivas:** justificam por que uma instância foi classificada de certo modo
- **Contrastivas:** explicam por que não foi classificada como esperado (*Por que NÃO?*)
- Foco do artigo: definir explicações contrastivas mais gerais e úteis

Preliminares

Conceitos básicos:

- *Atributos: booleanos, categóricos ou numéricos.*
- *Classificador binário: mapeia instância $\rightarrow \{0,1\}$.*
- *Árvores de decisão:*
 - *cada nó = condição;*
 - *cada folha = classe/resultado*
- *Random Forests \rightarrow conjunto de árvores \rightarrow voto majoritário.*



Melhorando a generalidade

- Explicações podem ser muito específicas ou vagas.
- Solução: reescrever instâncias em condições booleanas.

RENDA = 18K

EMPRÉSTIMO =
NÃO PAGO

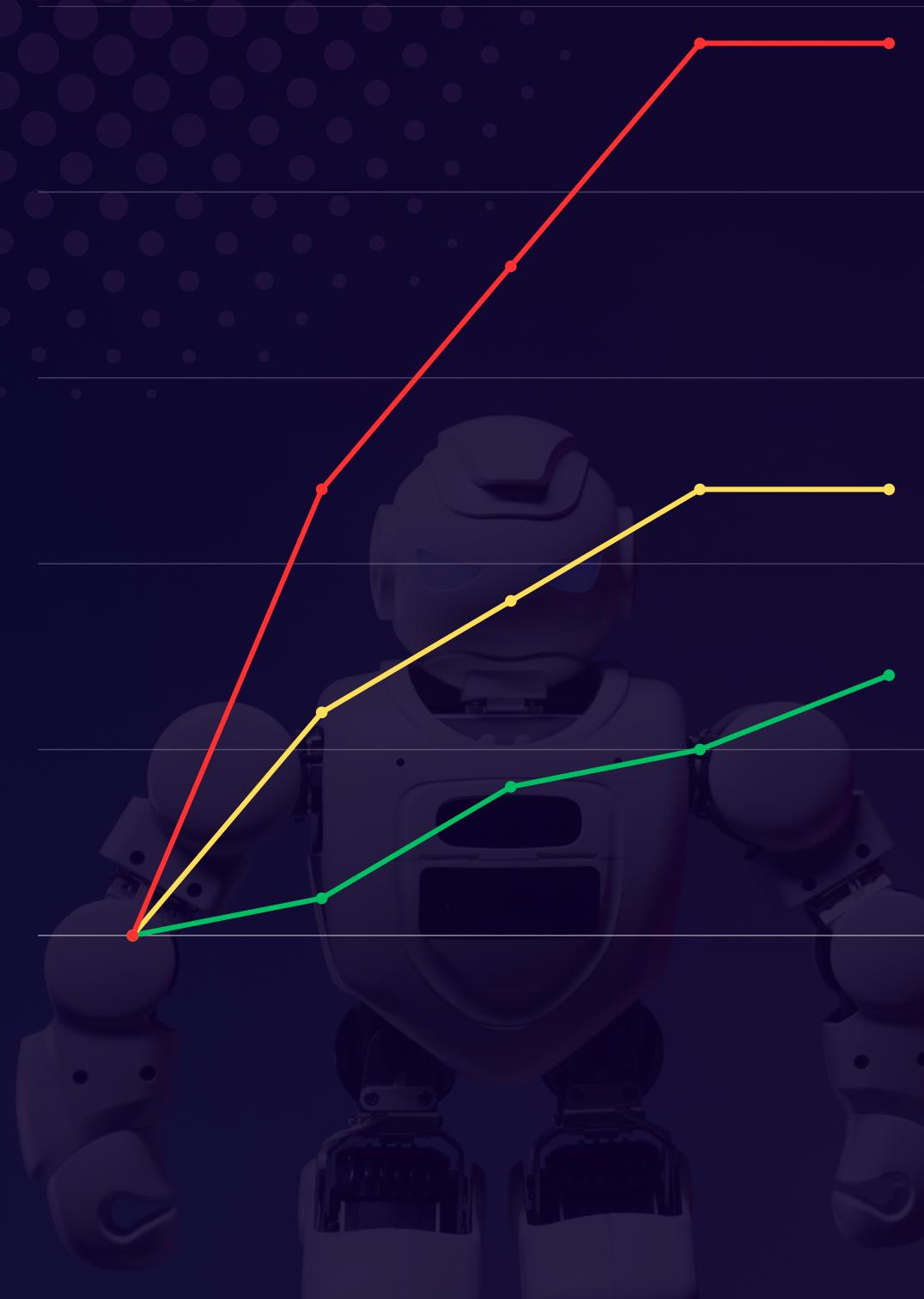


(RENDA \geq 20K)
(RENDA \geq 30K)
(EMPRÉSTIMO PAGO)



Explicações Contrastivas

Tipo	Definição	Exemplo
Simples	muda apenas um atributo para inverter a decisão.	Modelo classifica “ Negar Empréstimo ” porque $Renda = 18k$. Se fosse $Renda \geq 20k \rightarrow$ decisão mudaria para “ Aprovar ”.
Subconjunto-mínimo	menor conjunto de atributos suficientes para mudar a decisão (pode ser mais de um).	Decisão: “ Negar ”. Mudar apenas “Renda” não basta. Mas mudar {“Renda $\geq 20k$ ” e “Idade ≥ 25 ”} já inverte para “ Aprovar ”.
Tamanho mínimo	altera o menor número de atributos possíveis (pode não ser único).	Modelo usa 5 atributos. Alterar só 1 atributo (“Renda $\geq 30k$ ”) já muda para “ Aprovar ”. Então essa é a explicação de tamanho mínimo.



Complexidade Computacional

- Nem sempre existe explicação contrastiva.
- Reconhecer contrastiva = P (fácil).
- Subconjunto mínimo = coNP-completo (difícil).
- Abductiva ainda mais complexa.

Contribuições

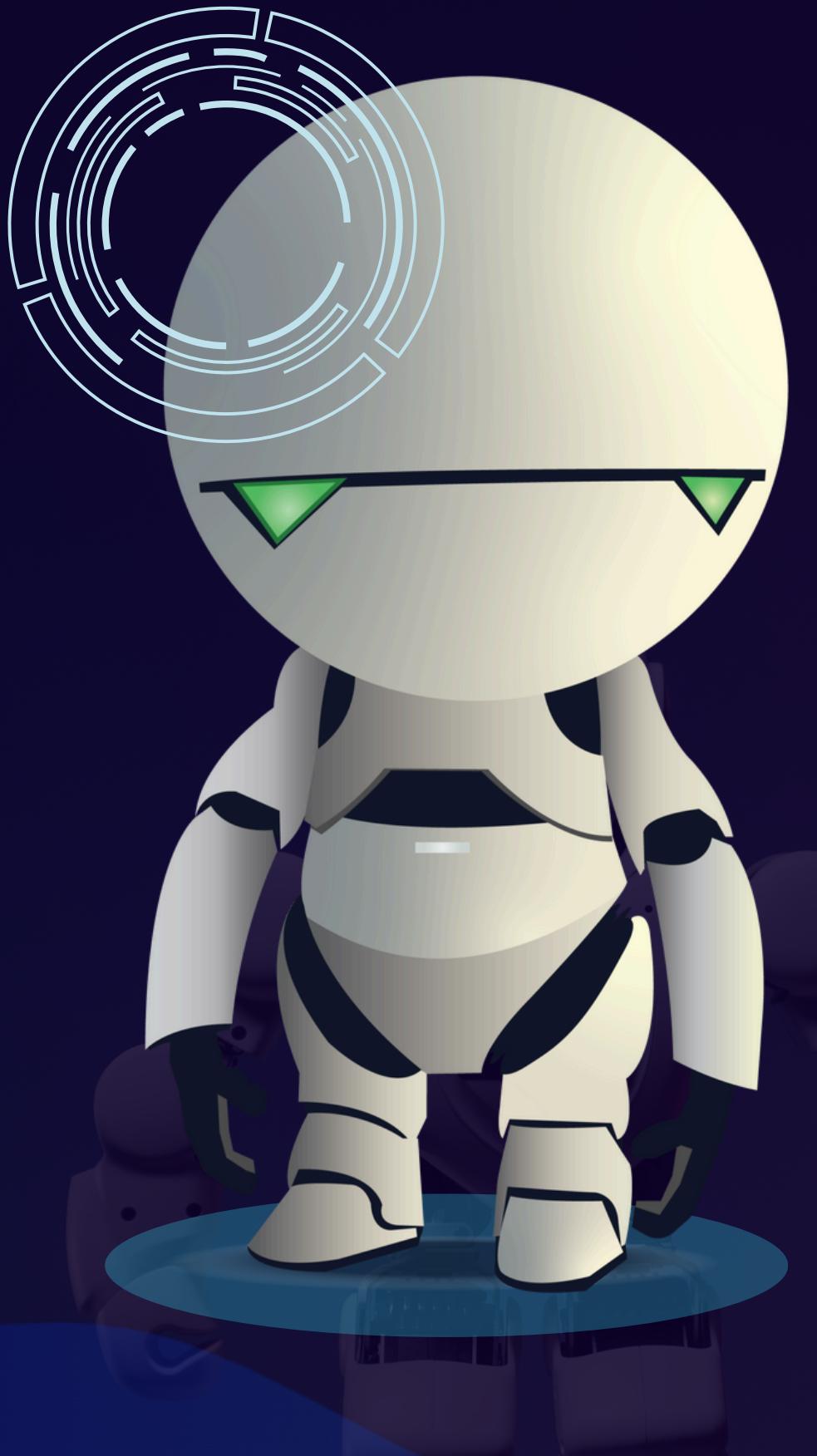
- *Definição de explicações contrastivas para classificadores baseados em árvores*
- *Análise da complexidade computacional para reconhecê-las*
- *Proposta de algoritmo para gerar explicações mínimas usando MAXSAT*
- *Avaliação experimental em 20 conjuntos de dados*



Metodologia

- Algoritmo baseado em **Partial MaxSAT**
- **Classificadores:** Random Forests (Scikit-Learn, 100 árvores)
- **Dados:** 20 datasets (Kaggle, OpenML, UCI)
- **Técnica:** PyXAI + MAXSAT solver
- **Métricas:**
 - Tempo de execução
 - Tamanho da explicação
 - Sucesso em encontrar explicações mínimas



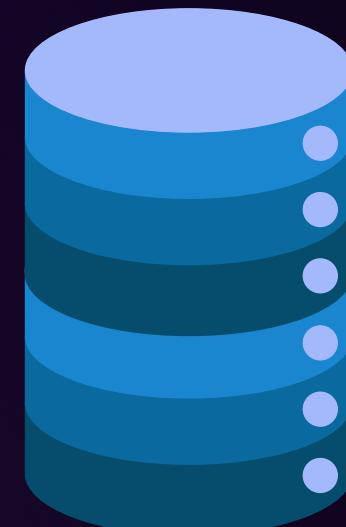


Resultados

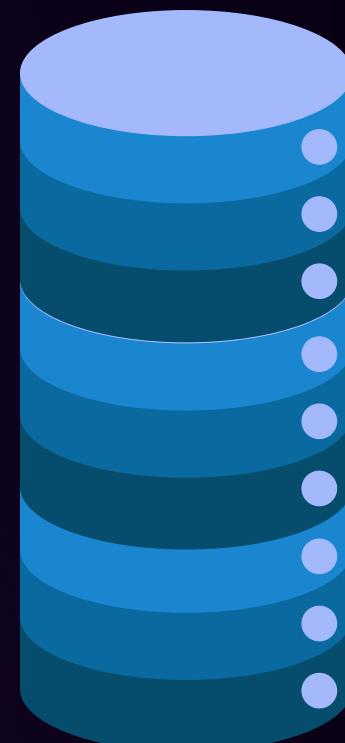
- *Algoritmo obtém explicações mínimas na maioria dos casos*
- *Tempo médio: ~13,5s por instância em datasets sem timeout*
- *Problemas em datasets muito grandes*
 - (exemplos: MNIST, Gisette, Dexter, Kaggle, UCI, OpenML)
- *Alterar em média 12,8% dos atributos já muda a classificação*



Balance-scale,
BreastTumor,
German



Bank,
Spambase,
Melb



Adult,
Default-
payment,
MNIST38,
Gisette

Comparação com Trabalhos Relacionados

- Abordagens existentes focam em instâncias contrastivas próximas
- Problemas: dependem de distâncias, escalas e heurísticas
On Contrastive Explanations for...
- Proposta do artigo:
 - Mais geral (conjunto de instâncias, não apenas uma)
 - Mais robusta
 - Evita suposições arbitrárias sobre métricas

Conclusão

- *Explicações contrastivas para modelos de árvores são:*
 - *Viáveis computacionalmente.*
 - *Mais gerais e informativas que abordagens clássicas.*
- *Algoritmo baseado em MAXSAT funciona bem em prática, exceto em datasets massivos.*
- *Contribui para IA explicável (XAI) confiável.*



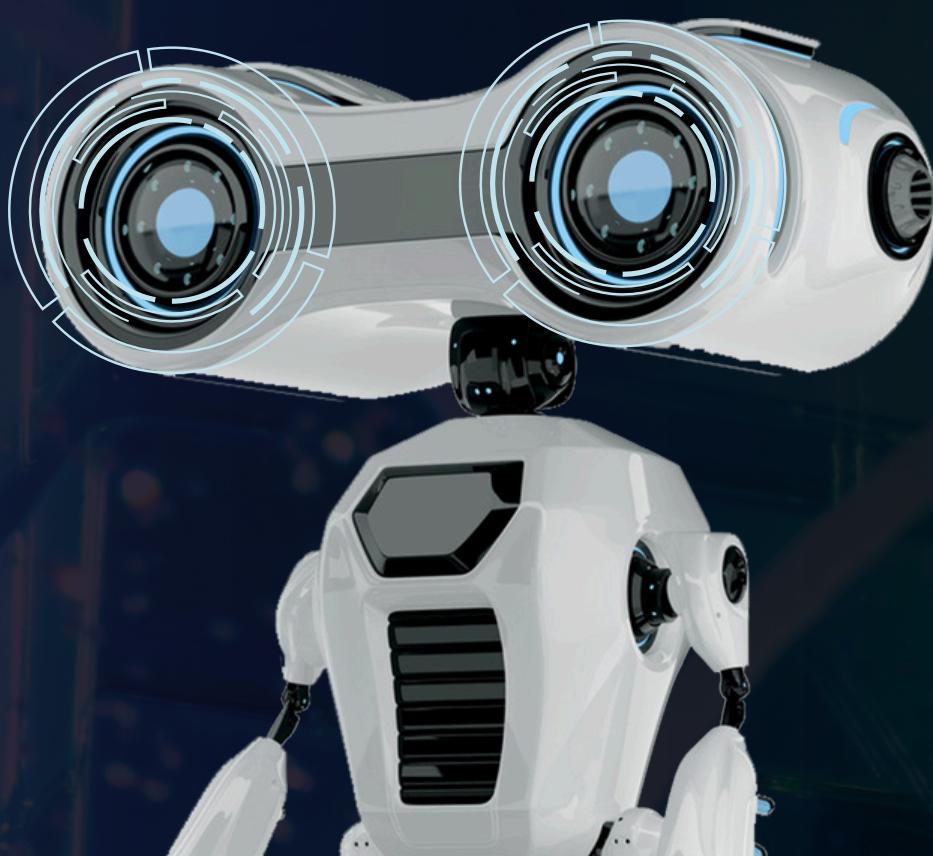


Trabalhos Futuros

- *Otimizar escalabilidade para datasets grandes*
- *Explorar visualizações mais comprehensíveis para humanos*
- *Estender a outros tipos de modelos além de Random Forests*

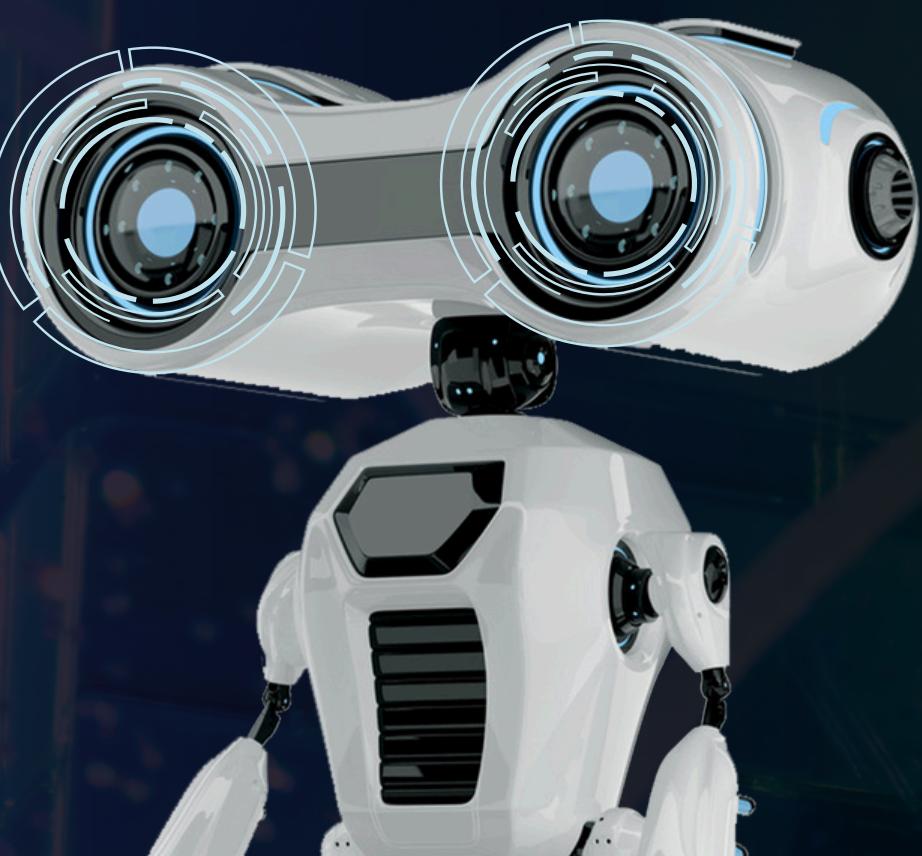
Perguntas

- *O que já existia antes do artigo em questão?*
- *Qual(is) problema(s) o artigo se propõe a resolver? E qual a novidade com relação aos trabalhos que já existiam?*
- *Qual(is) método(s) (definições, algoritmo, protocolo, ferramenta, modelagem, demonstrações) foram desenvolvidos e/ou usados?*
- *Qual(is) resultado(s) foram obtidos?*



Perguntas

- Qual(is) seriam as limitações do trabalho?
- Qual(is) próximos passos a serem desenvolvidos?
- Qual(is) conceitos do artigo você considera mais difíceis ou complexos?
- Que partes do conteúdo da disciplina foram utilizados no artigo?



Obrigado!

Explicações Contrastivas em Classificadores Baseados em Árvores

Acabou né? xD