

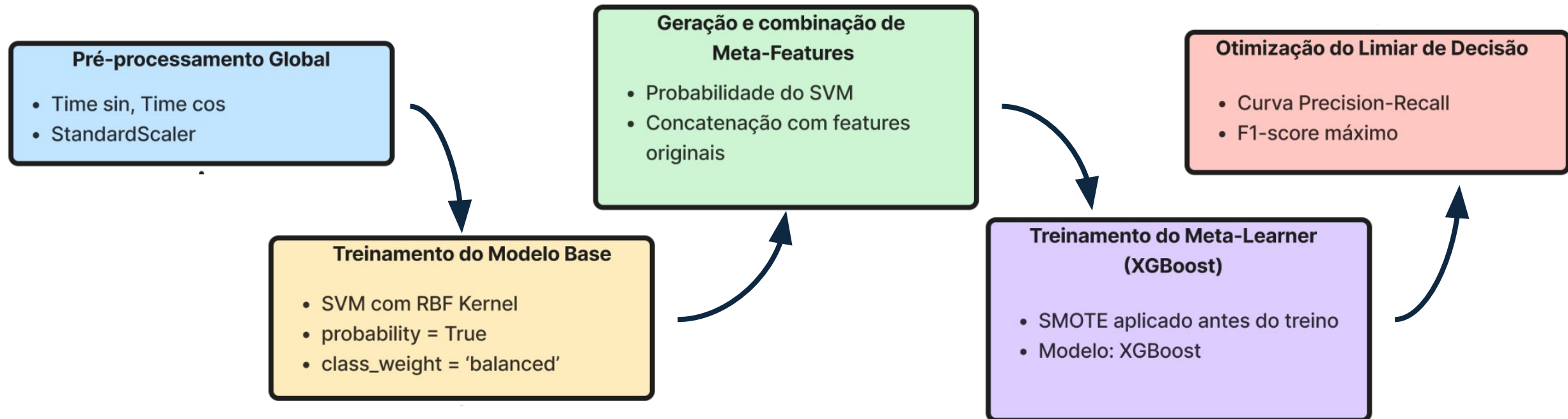
Motivação (Aplicação e Computação)

- **Impacto da Fraude (Aplicação):**
 - Perdas globais anuais estimadas em mais de US\$ 5 trilhões, de acordo com a *Association of Certified Fraud Examiners*.
 - Prejuízo médio de US\$ 1,7 milhão por caso entre janeiro de 2022 e setembro de 2023.
 - Impacto direto em instituições financeiras, comerciantes e na confiança do consumidor.
- **Desafio Computacional (Computação):**
 - **Problema Central:** Datasets de fraude são extremamente desbalanceados (apenas 0,172% de fraudes no nosso caso).
 - **Lacuna na Literatura:** Algoritmos tradicionais falham em detectar a classe minoritária. Faltam sistemas que refinem previsões de baixa confiança e corrijam erros de um classificador inicial.
 - **Nossa Proposta:** Um sistema em cascata para resolver essa lacuna. Implementar um sistema que diminui a chance de clientes serem lesados por ações fraudulentas de terceiros.

Sistema híbrido

- **Arquitetura em Cascata (Stacking):**

- **Modelo Base (SVM):** Um SVM com *kernel RBF* atua como primeiro classificador. Ele gera probabilidades de uma transação ser fraude.
- **Meta-Learner (XGBoost):** Um modelo XGBoost que recebe as features originais MAIS as probabilidades do SVM para refinar a decisão final.



Protocolo Experimental

- **Dataset:**

- Uma base de dados: "*Credit Card Fraud Detection*" do Kaggle.
- 284.807 transações, com 492 fraudes (0,172%).

- **Condução dos Experimentos:**

- 10 execuções independentes para garantir robustez estatística.
- Divisão Estratificada: 70% para treino, 15% para validação e 15% para teste.

- **Otimização e Testes:**

- **Otimização de Parâmetros:** Parâmetros definidos com base em experimentação prévia. O limiar de decisão do nosso modelo foi otimizado na etapa de validação para maximizar o *F1-Score*.
- **Testes de Hipótese Usados:** Teste T Pareado, *Wilcoxon Signed-Rank* e *McNemar*.
- **Como foram usados:** Para comparar nosso modelo com os de referência e verificar se as diferenças de desempenho eram estatisticamente significativas, com um nível de significância de 0,05.

Resultados

- **Comparação com a Literatura:**

- **Artigo 1 (*Ensemble*):** *VotingClassifier* com 3 modelos (*Multilayer Perceptron*, *XGBoost* e *Logistic Regression*). Testado com dados desbalanceados e balanceados com *SMOTE*.
- **Artigo 2 (Híbrido *RF-AdaBoost*):** *RandomForest* gera features para um *AdaBoost*.
- **Artigo 3 (*IHT-Ensemble*):** *VotingClassifier* com 4 modelos (*Decision Tree*, *Random Forest*, *K-Nearest Neighbors* e *Multilayer Perceptron*) e balanceamento complexo (*IHT*).

- **Análise dos Resultados:**

- **Resultados de Desempenho (Tabela 1):** Nosso modelo ($F1=0,75$) foi superado no *F1-Score* pelos modelos dos Artigos 1 e 2 ($F1\approx0,84$). Porém, nosso *AUC-ROC* (0,95) foi o segundo melhor, mostrando alta capacidade discriminatória.
- **Resultados de Hipótese (Tabela 2):** Os testes confirmaram que as vitórias dos Artigos 1 e 2 em *F1-Score* foram estatisticamente significativas ($p < 0,05$). Não houve diferença significativa contra o Artigo 1 Balanceado em *F1-Score* e contra o Artigo 3 em *AUC-ROC*.

TABLE I
TABELA 1: MÉDIAS E DESVIOS PADRÃO DAS MÉTRICAS DE DESEMPENHO (10 EXECUÇÕES)

Modelo de Referência	Acurácia	Precisão	Recall	F1-Score	AUC-ROC
Método Proposto	0,9992 ± 0,0001	0,8253 ± 0,0663	0,6946 ± 0,0571	0,7513 ± 0,0368	0,9497 ± 0,0238
Artigo1_Desbalanceado	0,9995 ± 0,0001	0,9471 ± 0,0289	0,7541 ± 0,0431	0,8393 ± 0,0352	0,8770 ± 0,0215
Artigo1_Balanceado_SMOTE	0,9992 ± 0,0001	0,7556 ± 0,0477	0,8230 ± 0,0293	0,7870 ± 0,0315	0,9113 ± 0,0146
Artigo2_RF_AdaBoost	0,9995 ± 0,0001	0,9221 ± 0,0195	0,7649 ± 0,0383	0,8356 ± 0,0244	0,8824 ± 0,0192
Artigo3_IHT_Ensemble	0,0245 ± 0,0074	0,0018 ± 0,0000	1,0000 ± 0,0000	0,0035 ± 0,0000	0,9554 ± 0,0168

TABLE II
TABELA 2: P-VALORES DOS TESTES DE HIPÓTESE (MODELO PROPOSTO VS. MODELOS DE REFERÊNCIA)

Modelo Proposto vs Modelo de Referência	Métrica	Teste T Pareado (p-value)	Wilcoxon Signed-Rank (p-value)
Artigo1_Desbalanceado	Acurácia	0,0002	0,0020
	Precisão	0,0005	0,0020
	Recall	0,0140	0,0312
	F1-Score	0,0002	0,0020
	AUC-ROC	0,0000	0,0020
Artigo1_Balanceado_SMOTE	Acurácia	0,7474	0,6875
	Precisão	0,0198	0,0273
	Recall	0,0000	0,0020
	F1-Score	0,0628	0,0645
	AUC-ROC	0,0006	0,0039
Artigo2_RF_AdaBoost	Acurácia	0,0001	0,0020
	Precisão	0,0034	0,0020
	Recall	0,0018	0,0039
	F1-Score	0,0000	0,0020
	AUC-ROC	0,0000	0,0020
Artigo3_IHT_Ensemble	Acurácia	0,0000	0,0020
	Precisão	0,0000	0,0020
	Recall	0,0000	0,0020
	F1-Score	0,0000	0,0020
	AUC-ROC	0,5948	1,0000

Conclusão

- **Objetivos Atingidos:**
 - **Proposta de um Novo Sistema:** Propusemos e implementamos com sucesso um sistema híbrido em cascata *SVM-XGBoost*, com código aberto.
 - **Validação Robusta:** O sistema foi rigorosamente avaliado contra três abordagens de referência da literatura usando um protocolo de 10 execuções e testes estatísticos.
 - **Desempenho Competitivo:** Embora não tenha superado os melhores modelos em *F1-Score*, o método proposto alcançou um desempenho comparável e competitivo, oferecendo um excelente poder discriminatório (alto *AUC-ROC*).
- **Próximos Passos:**
 - Otimização de hiperparâmetros dos modelos base e *meta-learner*.
 - Exploração de outros modelos na arquitetura em cascata.