



Autor: Vitor Rodrigues Izidoro  
Orientador: Prof. Fabricio Enembreck

1. Introdução

- Mineração de fluxo de dados em tempo real
- Redução de complexidade com Seleção de Instâncias
- Impacto em algoritmos ensembles (ARF, SRP)

2. Objetivo(s)

Avaliar e desenvolver técnicas de seleção de instâncias para ensembles orientados a fluxos de dados, viabilizando a aplicação desses algoritmos em cenários de larga escala.

3. Metodologia

Figura 01 – Mapa mental da metodologia

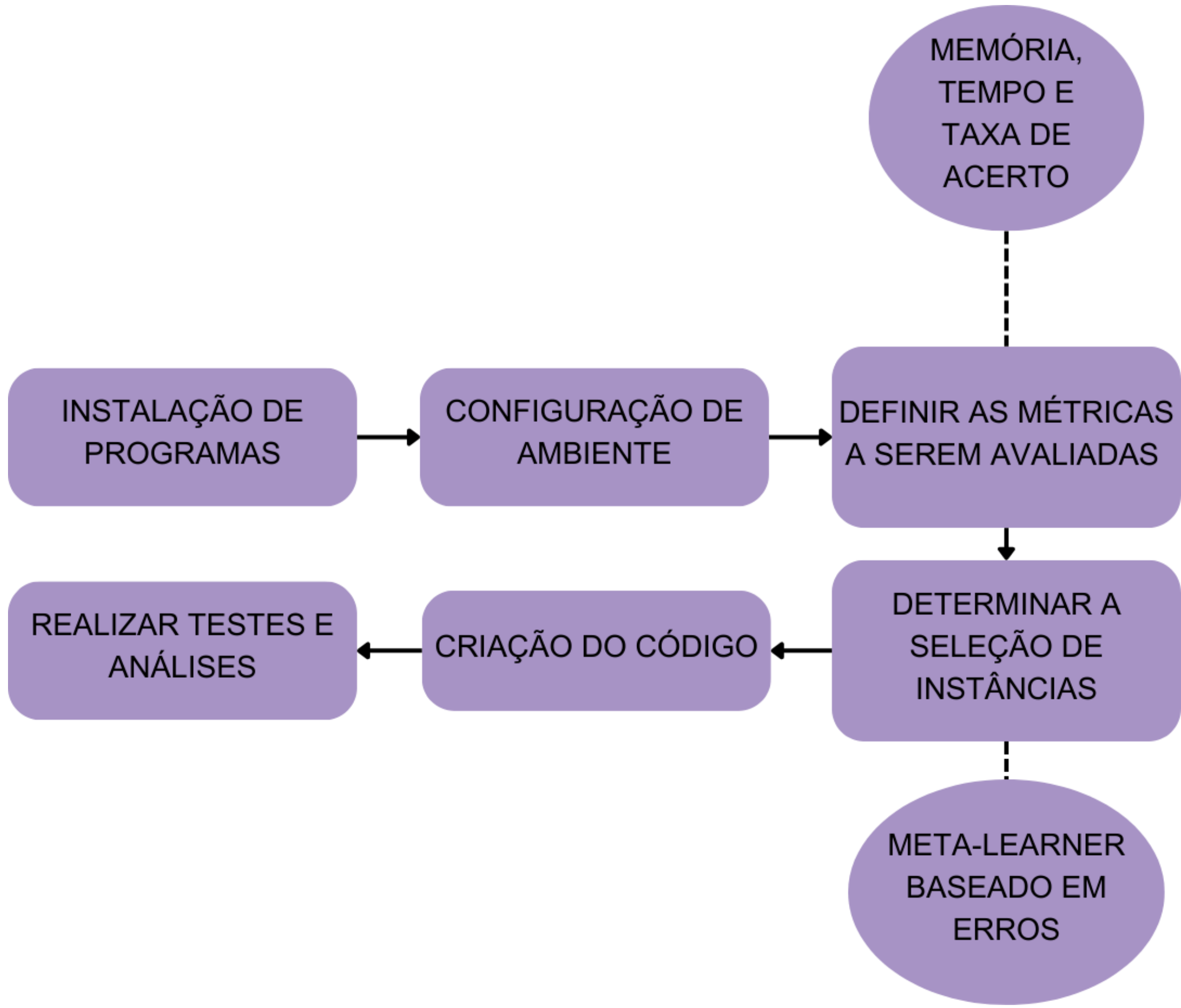


Tabela 01 – Pseudocódigo

```
sadClasse InstanceSelectedClassifier
variáveis:
    classifier: classificador atual;
Função resetLearningImpl():
    Inicializa 'classifiers' como novo classificador;
    Reseta aprendizagem de 'classifier';
Função trainOnInstanceImpl(instância):
    Se classe verdadeira da instância igual classe prevista
    pelo classificador:
        Não faça nada;
    Senão:
        Treine o classificador na instância;
Função getVotesForInstance(instância):
    Retorna os votos do classificador para a instância;
```

O código principal para este projeto foi desenvolvido utilizando como inspiração o código do DriftDetectionMethodClassifier, .java pois a estrutura de seleção do learner, escolha de arff's externos já havia sido desenvolvida. A ideia base para a criação do código era que tal seleção de instâncias teria como objetivo reduzir a quantidade de instâncias que seriam utilizadas no treinamento e aumentar a taxa de acerto dos Algoritmos.

“AVALIAÇÃO DO IMPACTO DE TÉCNICAS DE SELEÇÃO DE INSTÂNCIAS EM ENSEMBLES ORIENTADOS A FLUXOS DE DADOS”

Curso: bacharelado em ciência da computação.  
Escola: Politécnica.  
Campus: PUCPR.

4. Discussão e Resultado(s)

Nos casos testados, a seleção de instâncias sacrifica o percentual de acerto em troca de uma melhoria no custo computacional e na velocidade de treinamento. A escolha de realizar a seleção de instâncias depende muito da necessidade de cada projeto: se for necessário um processamento mais rápido e com menor custo computacional, a seleção de instâncias seria uma solução viável.

Tabela 02 – Tabela de ranking da taxa de acerto

Dataset	Levering	IS-levering	ARF	IS-ARF	OzaBoost	IS-OzaBoost	SRP	IS-SRP
Airlines	4	8	2	6	3	7	1	5
Kdd99	7	6	4	3	8	5	2	1
Kddcup	5	4	3	8	7	6	1	2
Keystroke	5	8	2	4	6	7	1	3
Luxembourg	4	6	1	2	7	8	3	5
NOAA	3	5	1	6	4	8	2	7
Nomao	6	5	2	3	8	7	1	4
Outdoor	7	5	4	3	8	6	2	1
Ozone	2	5	1	7	4	8	3	6
Poker	1	4	6	3	7	8	5	2
Rialto	6	5	4	3	8	7	1	2
Média:	4,545	5,54	2,72	4,36	6,36	7,00	2,00	3,45

Tabela 03 – Tabela de ranking de tempo

Dataset	levering	IS-levering	ARF	IS-ARF	OzaBoost	IS-OzaBoost	SRP	IS-SRP
Airlines	7	5	6	3	2	1	8	4
Kdd99	2	3	5	7	1	4	8	6
Kddcup	6	7	4	3	1	2	8	5
Keystroke	6	3	7	4	2	1	8	5
luxembourg	7	3	6	4	2	1	8	5
NOAA	6	3	7	4	2	1	8	5
nomao	6	3	7	4	2	1	8	5
outdoor	7	3	6	4	1	2	8	5
ozone	7	3	6	4	2	1	8	5
poker	6	3	7	4	2	1	8	5
rialto	6	3	7	4	2	1	8	5
Média:	6	3,54	6,18	4,09	1,72	1,45	8	5

5. Conclusão ou Considerações finais

Os ensembles, apesar de serem precisos, são lentos e pesados para aplicações em larga escala. A seleção de instâncias, por sua vez, reduz o consumo de recursos computacionais, mas sacrifica a taxa de acerto. É necessário mais desenvolvimento em técnicas de seleção que minimizem essa perda de precisão, vendo que a simplicidade da técnica avaliada potencializou as limitações.

6. Referências

Heitor Murilo Gomes, Jean Paul Barddal, Fabrício Enembreck, and Albert Bifet. 2017. A Survey on Ensemble Learning for Data Stream Classification. ACM Comput. Surv. 50, 2, Article 23 (March 2017), 36 pages. DOI: <https://doi.org/10.1145/3054925>

BARDDAL, JEAN PAUL ; GOMES, HEITOR MURILO ; Enembreck, Fabrício . SNCStream. In: the 30th Annual ACM Symposium, 2015, Salamanca. Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15. v. 1. p. 935-940.

MOA. Home-Page do aplicativo MOA. Disponível em: <https://moa.cms.waikato.ac.nz/documentation/>

Eclipse Foundation. Home-Page da IDE Eclipse. Disponível em: <https://eclipseide.org/>

Nicoletti, Maria & Santos, Flávia. A Família de Algoritmos Instance-Based Learning (IBL). Universidade Federal de São Carlos (UFSCar).