



WPCCG'2019

MÉTODO DE EXTRAÇÃO DE ATRIBUTOS PARA CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO USANDO AUTOENCODERS

Rafael Fernandes Siqueira
Prof^a Dr^a Helyane Bronoski Borges

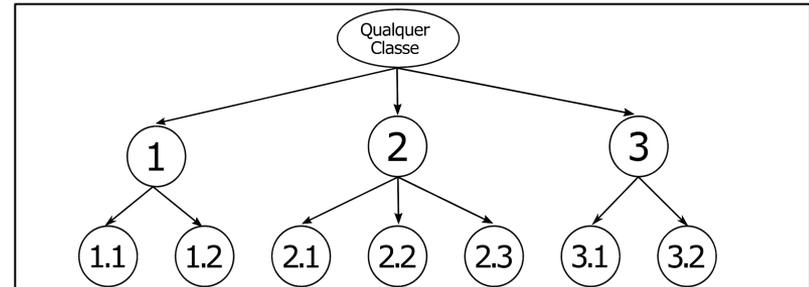
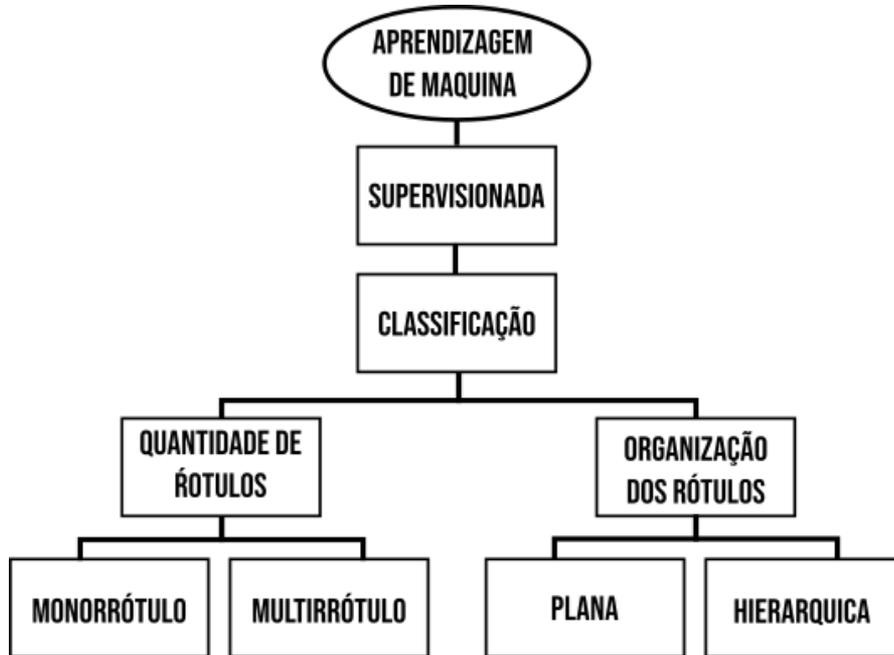


Ponta Grossa
2019

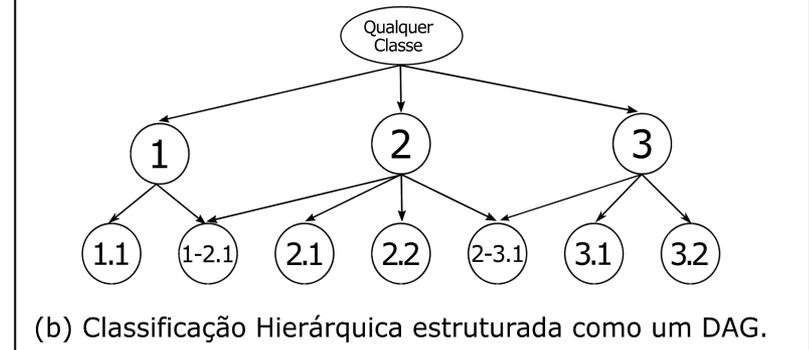
ROTEIRO

- Classificação de Dados
- Redução de Dimensionalidade
- Autoencoders
- FEAE-HMC
- Experimentos e Resultados
- Conclusão

CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO



(a) Classificação Hierárquica estruturada como uma árvore.



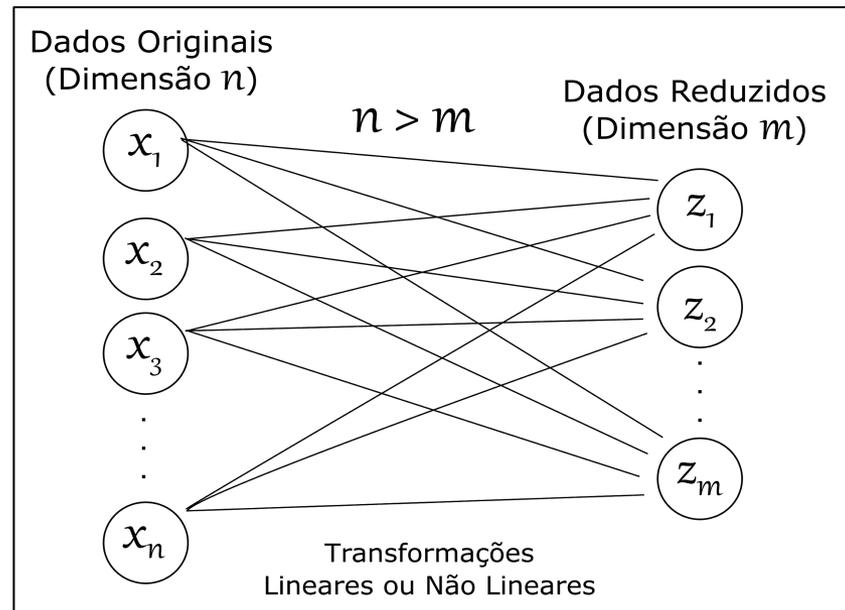
(b) Classificação Hierárquica estruturada como um DAG.

Fonte: Adaptado (FREITAS; CARVALHO, 2007)

- Predição de proteínas;

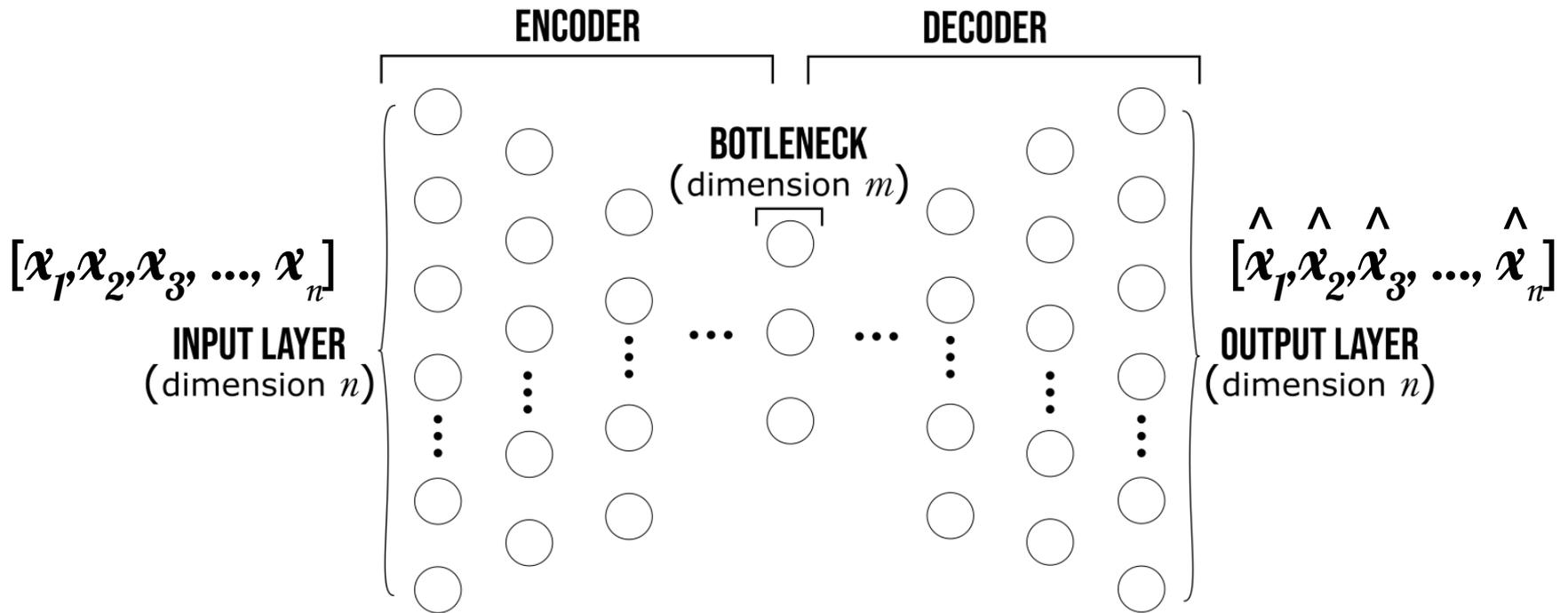
REDUÇÃO DE DIMENSIONALIDADE

- Problemas de capacidade computacional;
- Aumentar capacidade de generalização.
- Extração de Atributos:



Fonte: Autoria própria

AUTOENCODERS

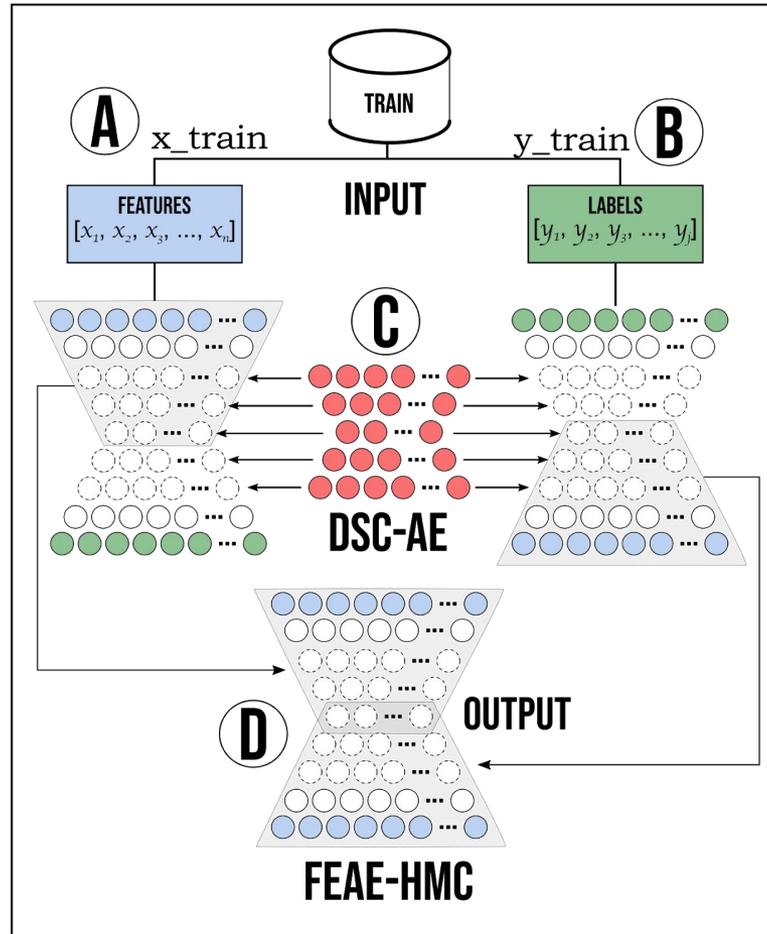


Fonte: Autoria própria

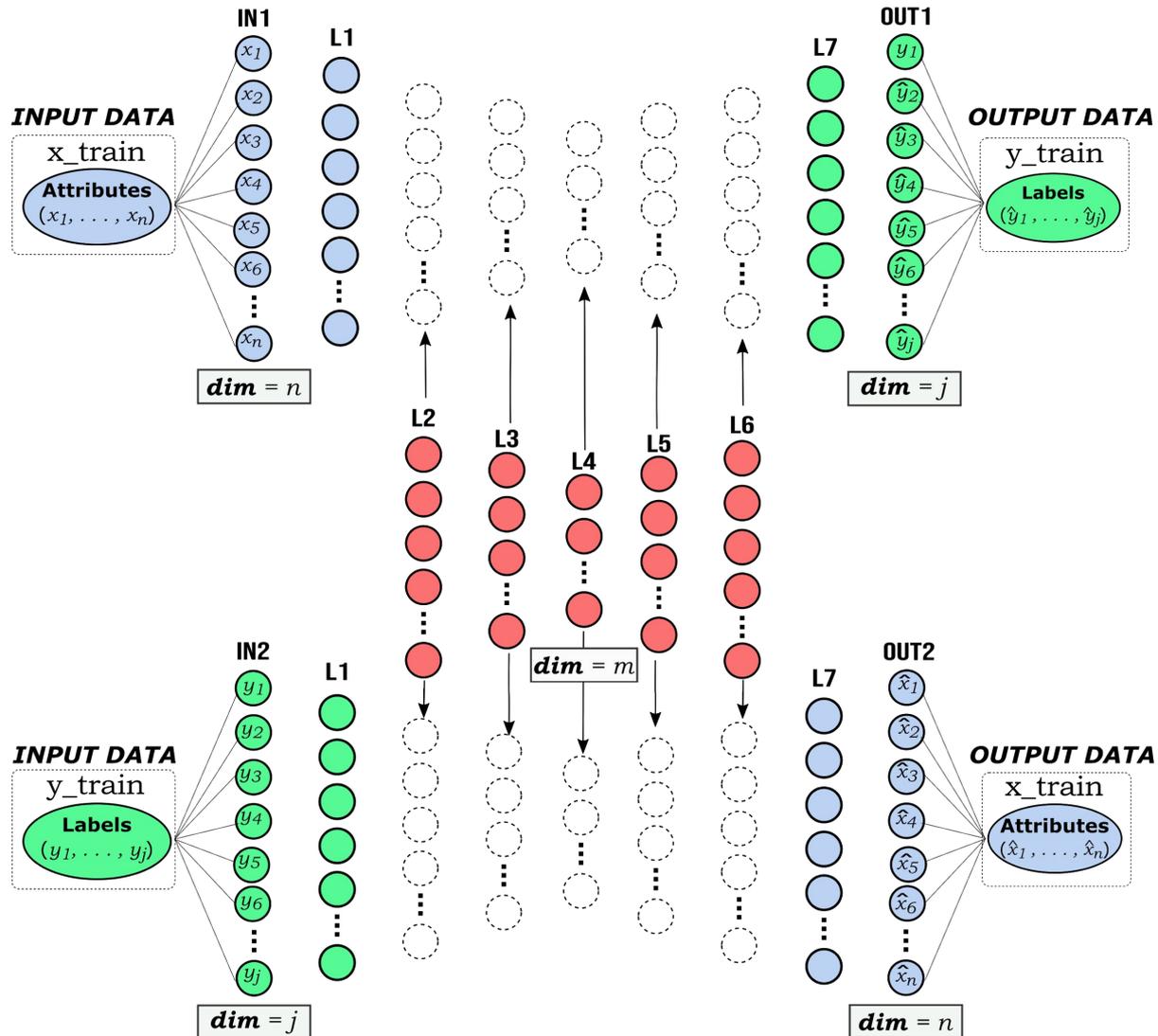
MÉTODO – FEAE-HMC

- *Feature Extraction based on Autoencoders for the Hierarchical Multi-label Classification.*
- Adaptações na rede Autoencoder tradicional;
 - Correlações entre os atributos e os rótulos.
- Desempenho preditivo equivalente ou superior a base original.

MÉTODO – FEAE-HMC

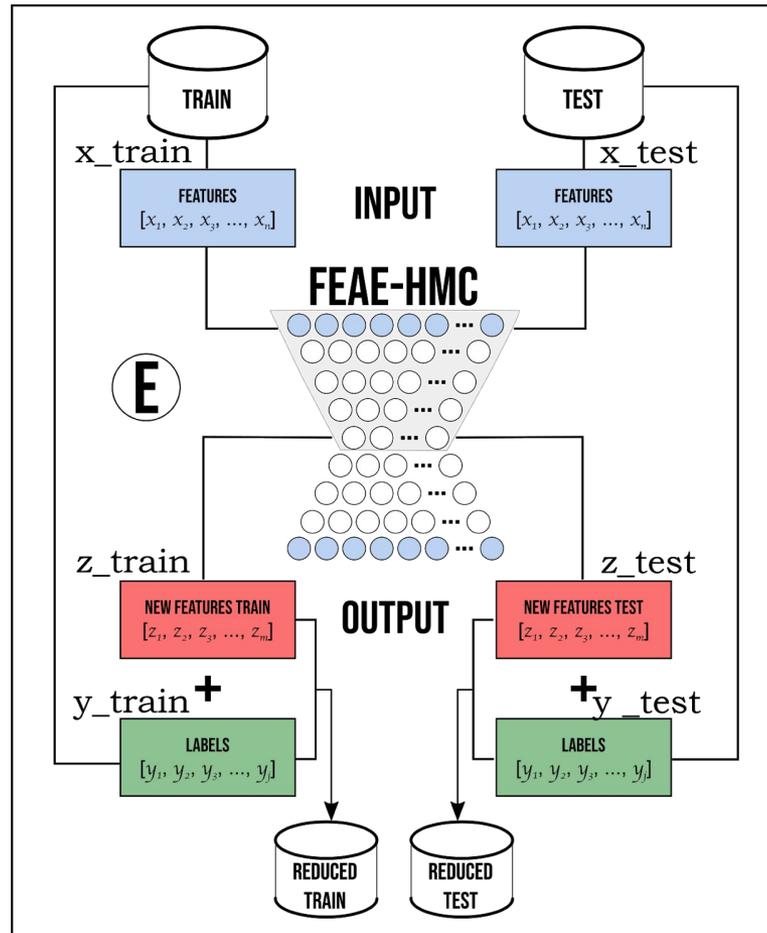


Fonte: Autoria própria



Fonte: Autoria própria

MÉTODO – FEAE-HMC



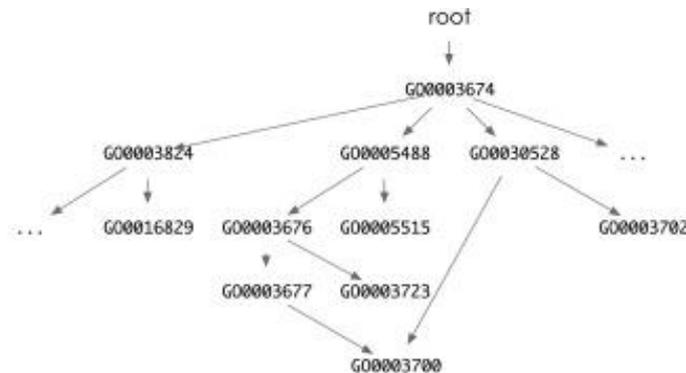
Fonte: Autoria própria

EXPERIMENTOS

- Bases de dados

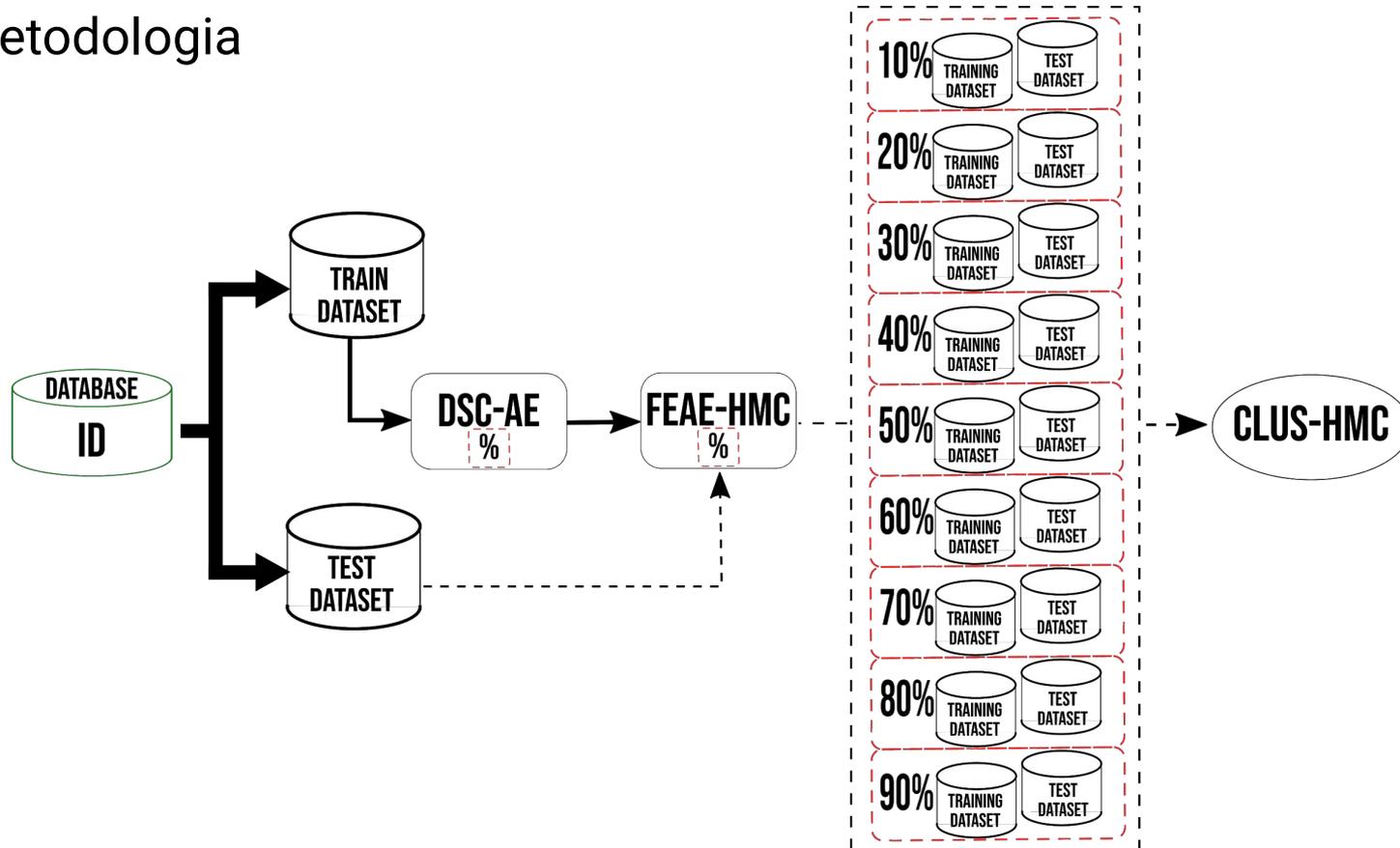
| ID | Nome | Amostras | Atributos | Classes | Classes/Amostra | Amostras/Classe |
|----|----------|----------|-----------|---------|-----------------|-----------------|
| 1 | Celcycle | 3751 | 77 | 4126 | 03/28 | 0/785 |
| 2 | Church | 3749 | 27 | 4126 | 03/28 | 0/786 |
| 3 | Derisi | 3719 | 63 | 4120 | 03/28 | 0/781 |
| 4 | Eisen | 2418 | 79 | 3574 | 03/28 | 0/492 |
| 5 | Expr | 3773 | 551 | 4132 | 03/28 | 0/789 |
| 6 | Gasch1 | 3758 | 173 | 4126 | 03/28 | 0/786 |
| 7 | Gasch2 | 3773 | 52 | 4132 | 03/28 | 0/789 |
| 8 | Pheno | 1586 | 69 | 3128 | 03/28 | 0/388 |
| 9 | Seq | 3900 | 478 | 4134 | 03/28 | 0/791 |
| 10 | Spo | 3697 | 80 | 4120 | 03/28 | 0/775 |

Fonte: Autoria própria



EXPERIMENTOS

- Metodologia



Fonte: Autoria própria

EXPERIMENTOS

- Exemplo: Comparativo - Dimensão Original x Bases Reduzidas

| ID | Atributos (<i>n</i>) | Novos atributos (<i>m</i>) conforme <i>Pr</i> | | | | | | | | |
|----|---------------------------|---|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| 1 | 77 | 69 | 61 | 53 | 46 | 38 | 30 | 23 | 15 | 7 |
| 2 | 27 | 24 | 21 | 18 | 16 | 13 | 10 | 8 | 5 | 2 |
| 3 | 63 | 56 | 50 | 44 | 37 | 31 | 25 | 18 | 12 | 6 |
| 4 | 79 | 71 | 63 | 55 | 47 | 39 | 31 | 23 | 15 | 7 |
| 5 | 551 | 495 | 440 | 385 | 330 | 275 | 220 | 165 | 110 | 55 |
| 6 | 173 | 155 | 138 | 121 | 103 | 86 | 69 | 51 | 34 | 17 |
| 7 | 52 | 46 | 41 | 36 | 31 | 26 | 20 | 15 | 10 | 5 |
| 8 | 69 | 62 | 55 | 48 | 41 | 34 | 27 | 20 | 13 | 6 |
| 9 | 478 | 430 | 382 | 334 | 286 | 239 | 191 | 143 | 95 | 47 |
| 10 | 80 | 72 | 64 | 56 | 48 | 40 | 32 | 24 | 16 | 8 |

Fonte: Autoria própria

RESULTADOS

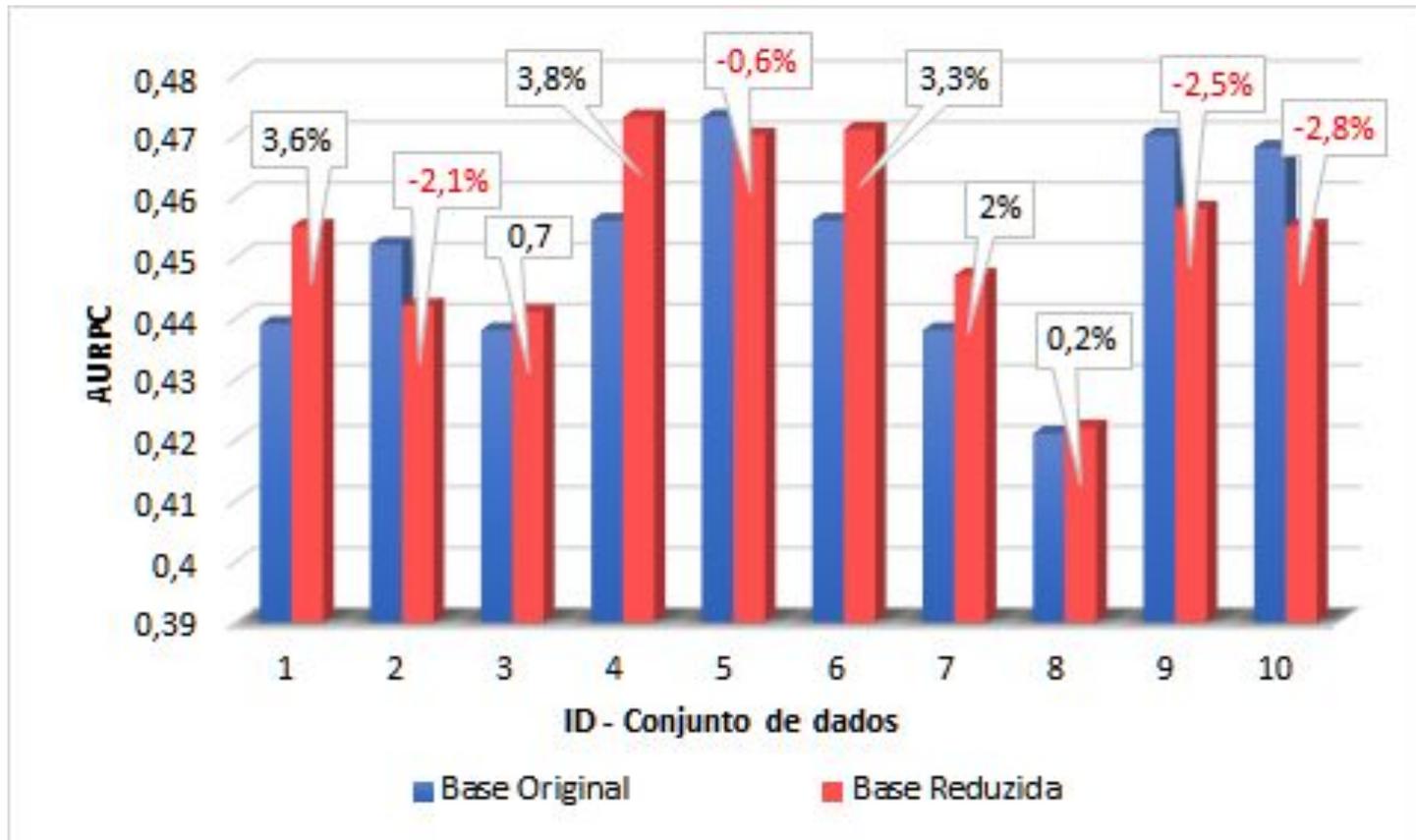
- Clus-HMC

| ID | TA | Percentual de Redução Aplicado | | | | | | | | | | Méd | Max | %Méd | %Max |
|----|-------|--------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | | | | | |
| 1 | 0,439 | 0,456 | 0,457 | 0,453 | 0,455 | 0,455 | 0,456 | 0,453 | 0,455 | 0,454 | 0,455 | 0,457 | 3,6% | 4,1% | |
| 2 | 0,452 | 0,441 | 0,441 | 0,446 | 0,444 | 0,443 | 0,445 | 0,442 | 0,439 | 0,440 | 0,442 | 0,446 | -2,1% | -1,3% | |
| 3 | 0,438 | 0,442 | 0,442 | 0,440 | 0,440 | 0,440 | 0,440 | 0,442 | 0,442 | 0,440 | 0,441 | 0,442 | 0,7% | 0,9% | |
| 4 | 0,456 | 0,470 | 0,475 | 0,473 | 0,469 | 0,472 | 0,476 | 0,475 | 0,474 | 0,476 | 0,473 | 0,476 | 3,8% | 4,4% | |
| 5 | 0,473 | 0,468 | 0,473 | 0,472 | 0,467 | 0,469 | 0,469 | 0,468 | 0,473 | 0,471 | 0,470 | 0,473 | -0,6% | 0,0% | |
| 6 | 0,456 | 0,474 | 0,468 | 0,472 | 0,471 | 0,469 | 0,472 | 0,468 | 0,472 | 0,472 | 0,471 | 0,474 | 3,3% | 3,9% | |
| 7 | 0,438 | 0,450 | 0,447 | 0,447 | 0,445 | 0,448 | 0,443 | 0,448 | 0,447 | 0,446 | 0,447 | 0,450 | 2,0% | 2,7% | |
| 8 | 0,421 | 0,421 | 0,421 | 0,421 | 0,422 | 0,422 | 0,423 | 0,422 | 0,422 | 0,422 | 0,422 | 0,423 | 0,2% | 0,5% | |
| 9 | 0,470 | 0,459 | 0,460 | 0,459 | 0,461 | 0,462 | 0,455 | 0,457 | 0,458 | 0,452 | 0,458 | 0,462 | -2,5% | -1,7% | |
| 10 | 0,468 | 0,454 | 0,455 | 0,455 | 0,457 | 0,457 | 0,452 | 0,457 | 0,451 | 0,458 | 0,455 | 0,458 | -2,8% | -2,1% | |

Fonte: Autoria própria

RESULTADOS

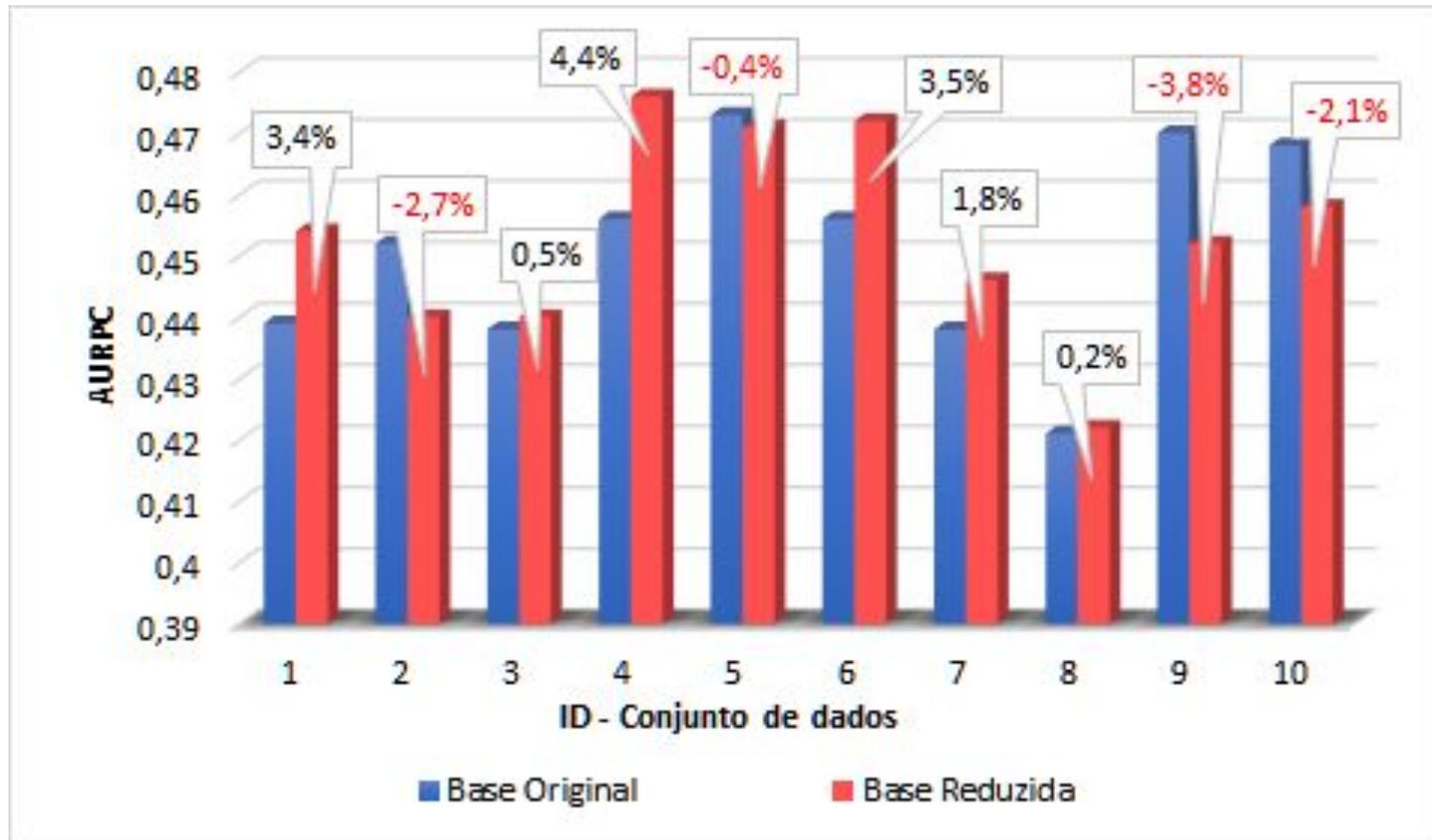
Dimensão Original x Média Bases Reduzidas



Fonte: Autoria própria

RESULTADOS

Dimensão Original x Dimensão 90% reduzida



Fonte: Autoria própria

CONCLUSÃO

- Novo método de extração de atributos FEAE-HMC;
- Redução de até 90% da dimensão original;
- Desempenho preditivo estatisticamente equivalentes;
- Trabalhos Futuros.

REFERÊNCIAS

BENGIO, Yoshua. Deep learning of representations for unsupervised and transfer learning. In: **Proceedings of ICML Workshop on Unsupervised and Transfer Learning**. [S.l.: s.n.], 2012. p. 17–36.

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation learning: A review and new perspectives. **IEEE transactions on pattern analysis and machine intelligence**, **IEEE**, v. 35, n. 8, p. 1798–1828, 2013.

BORGES, Helyane Bronoski. **Classificador hierárquico multirrótulo usando uma rede neural competitiva**. Tese (Doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba, Brasil, 2012.

FREITAS, Alex; CARVALHO, André. A tutorial on hierarchical classification with applications in bioinformatics. In: **Research and trends in data mining technologies and applications**. [S.l.]: IGI Global, 2007. p. 175–208.

MAATEN, Laurens Van Der; POSTMA, Eric; HERIK, Jaap Van den. Dimensionality reduction: a comparative. **J Mach Learn Res**, v. 10, n. 66-71, p. 13, 2009.