

# MATE33

## Tópicos em Inteligência Artificial III (Aprendizado Não Supervisionado)

**Trabalho final – Uma abordagem baseada em AM para o agrupamento dos dados de candidatos e vagas de trabalho disponíveis no site do MTE**

Dezembro, 2022

**Alunos: Anderson Boa Morte/Edmilson Jesus**

**Professor: Ricardo Rios**



**PGCOMP**  
Universidade Federal da Bahia

# Agenda

---



- 1. Contextualização do problema**
- 2. Visão Geral da Proposta**
- 3. Pré-processamento**
- 4. Aplicação dos modelos de AM**
- 5. Validação dos Resultados**
- 6. Considerações e trabalhos futuros**

# Contextualização do problema



- A taxa de desemprego no Brasil está em **8.7%** e atinge **9.5 milhões** de pessoas, no 3º trimestre de 2022 (IBGE, 2022).
- **SINE – Sistema Nacional de Emprego** vinculado ao Ministério do Trabalho, é um grande sistema público, responsável pela execução das políticas públicas de emprego no país.
- No SINE, atualmente, a busca de vagas por candidatos está **fortemente atrelada ao CBO** (Classificação Brasileira de Ocupações).

# Contextualização do problema



A seguir alguns números para ilustrar o cenário de contratações efetuadas via SINE:

- Somente **2.6%** das contratações do mercado de trabalho se dão através de atendimentos realizados via SINE.
- Em média são necessários **45 atendimentos** para que se efetive uma contratação via SINE.

# Contextualização do problema



**SINE aberto:** Iniciativa do Governo Federal em compartilhar dados dos trabalhadores cadastrados no SINE no intuito de aumentar a efetividade de contratações



O SINE Aberto é uma plataforma que visa aumentar a efetividade na colocação do trabalhador no mercado de trabalho, por meio do compartilhamento dos dados dos trabalhadores cadastrados no Sistema Nacional de Emprego (SINE).

Para isso, o governo federal propõe realizar parcerias com empresas, ampliando os atores que fazem intermediação de mão de obra e potencializando a geração de empregos.

# Contextualização do problema

---



Encontrar uma maneira de **agrupar os dados** sobre candidatos e vagas de forma a aumentar a taxa de sucesso no preenchimento das vagas de emprego.

# Agenda

---



1. Contextualização do problema
2. Visão Geral da Proposta
3. Pré-processamento
4. Aplicação dos modelos de AM
5. Validação dos Resultados
6. Considerações e trabalhos futuros



# Visão Geral da Proposta **etapa 1**



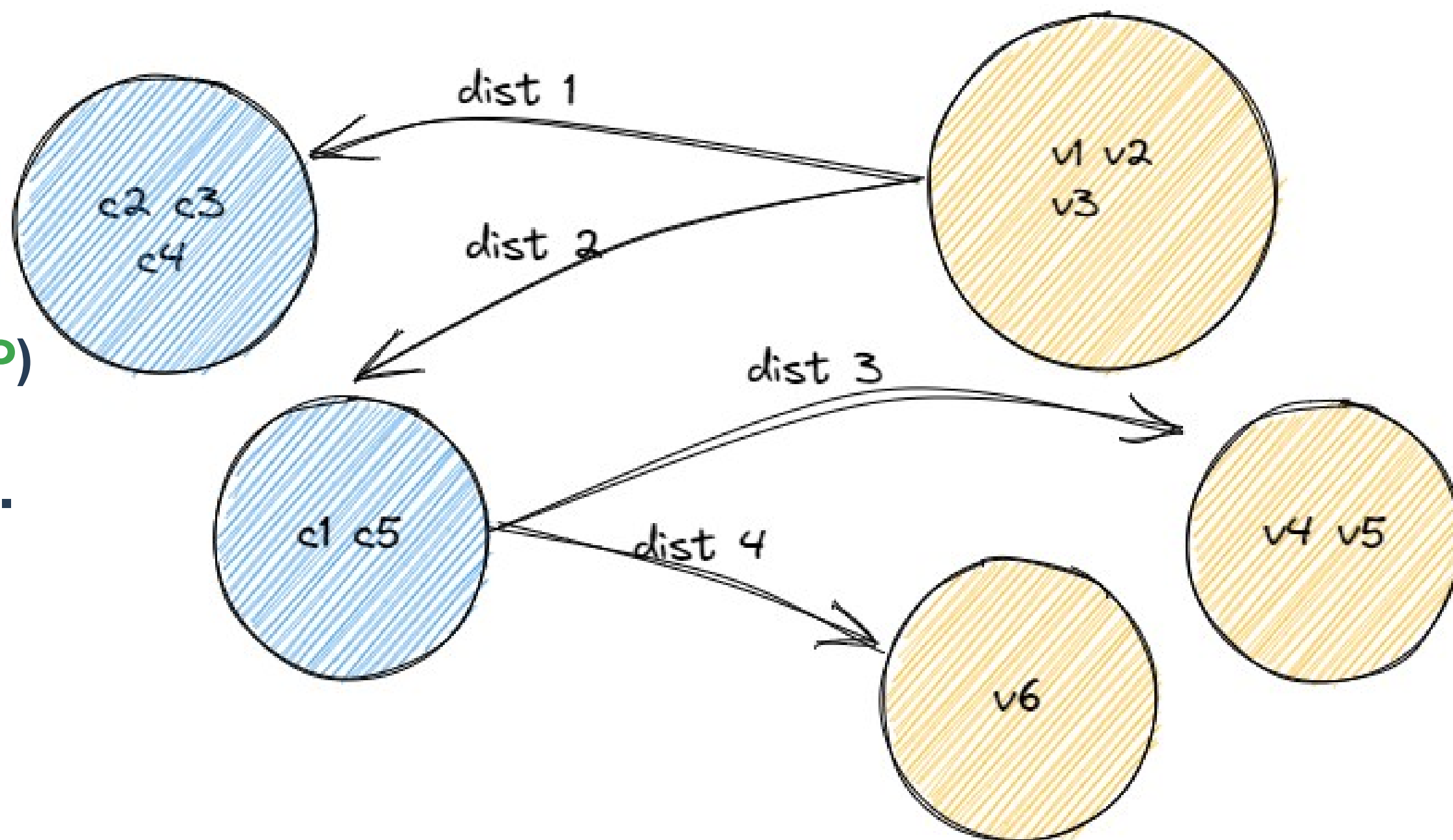
Avaliar algoritmos de agrupamento. Qual o algoritmo apresentou as melhores métricas?

Utilizar agrupamento particional e hierárquico p/:

- Avaliar medidas de similaridade/distância (**coseno**, **manhattan**, **euclidiana**)
- Avaliar índices de validação de agrupamento (**silhueta**, **DBI**, **GAP**)
- Avaliar métricas de integração (**single**, **average** e **complete-link**).
- Obter número ótimo de **clusters** para os algoritmos avaliados.

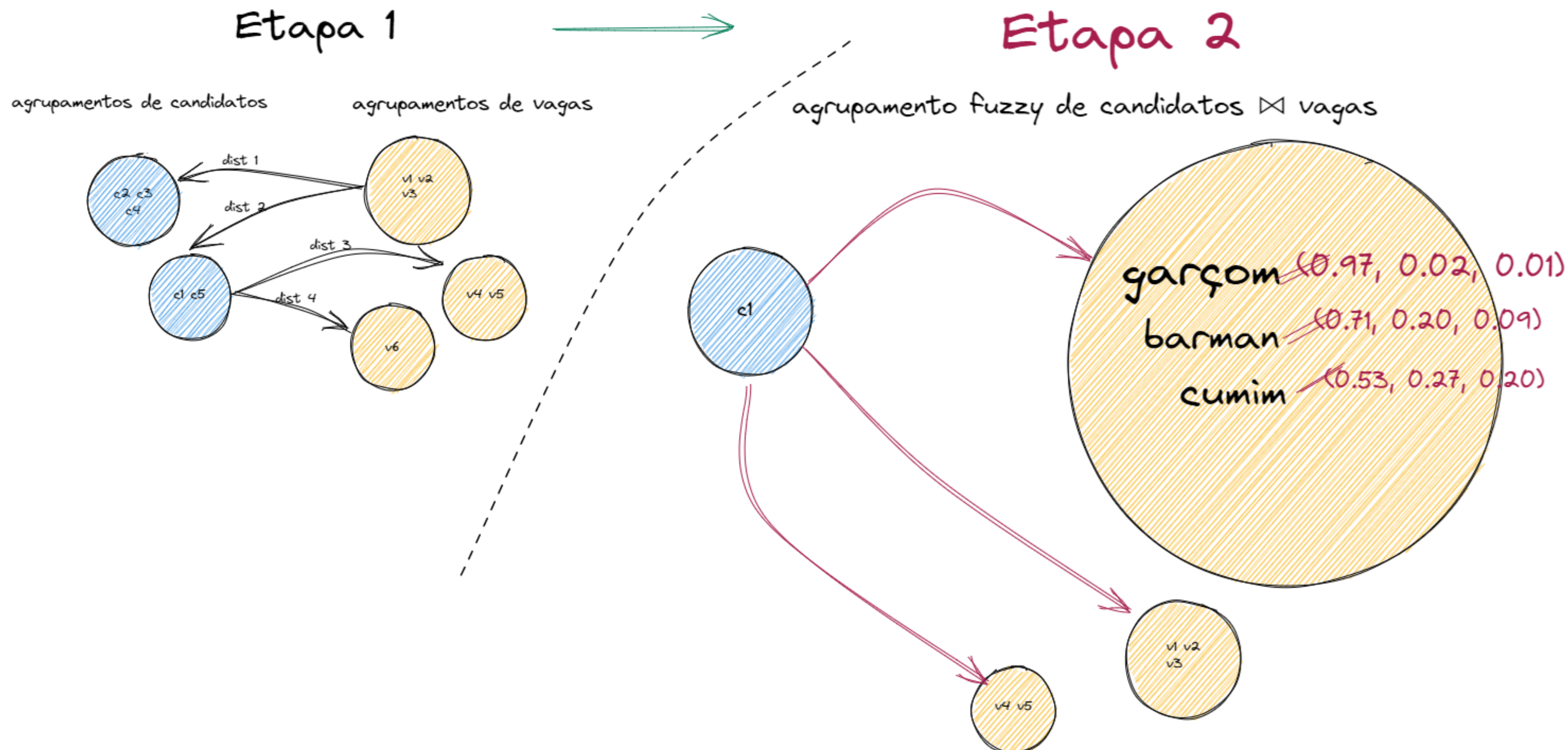
agrupamentos de candidatos

agrupamentos de vagas





# Visão Geral da Proposta **etapa 2**



# Visão Geral da Proposta **etapa 2**



## **Etapa 2:**

- Efetuar a junção (⋈) dos *datasets* de candidatos e vagas através de alguns campos Ex.: CBOs pretendidos e localidade.
- Submeter esse *dataset* único ao algoritmo de agrupamento *fuzzy* no intuito de determinar o grau de pertencimento do candidato a cada um dos *clusters* de vagas obtidos.

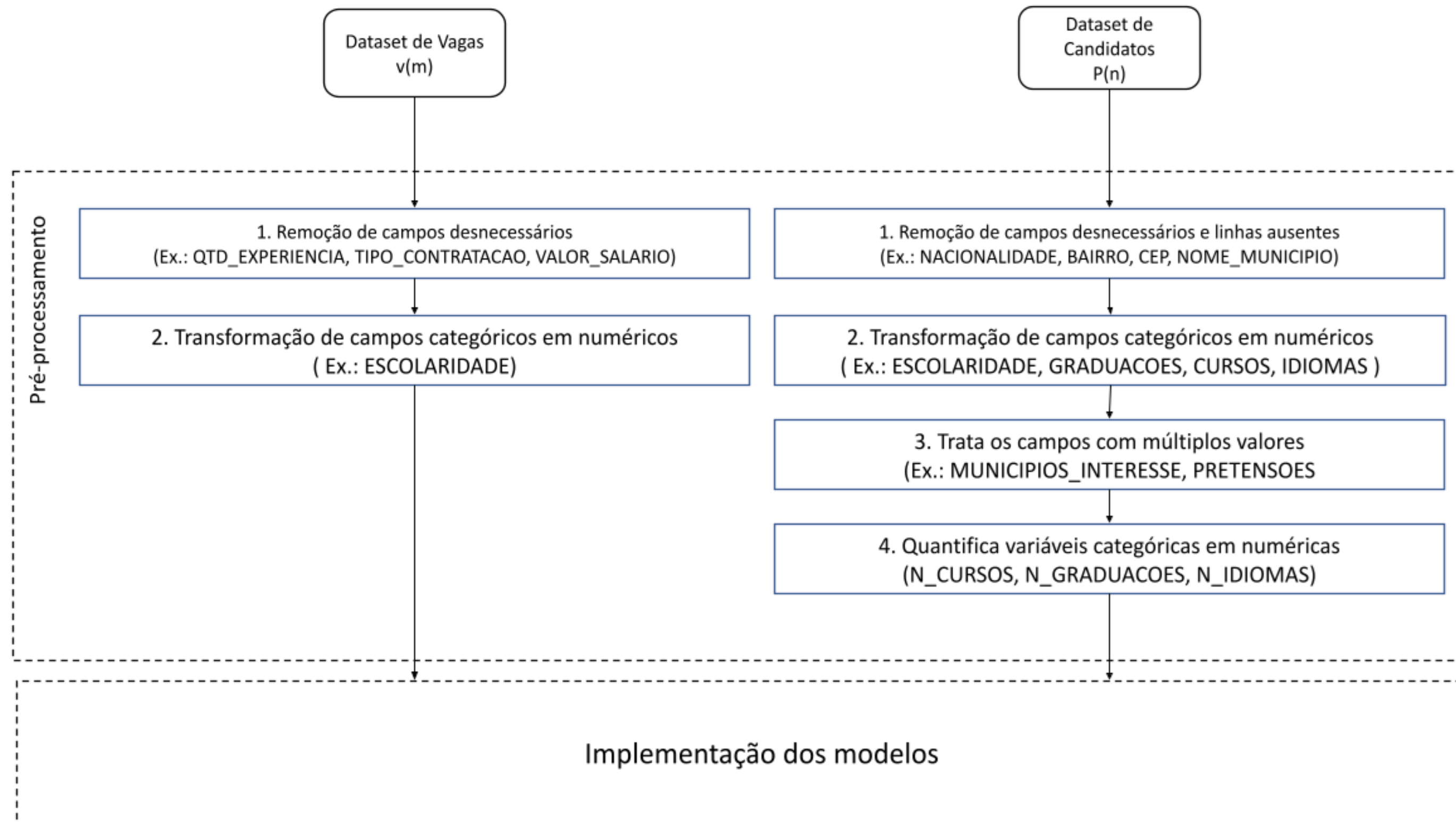
# Agenda

---



1. Contextualização do problema
2. Visão Geral da Proposta
3. Pré-processamento
4. Aplicação dos modelos de AM
5. Validação dos Resultados
6. Considerações e trabalhos futuros

# Pré-processamento





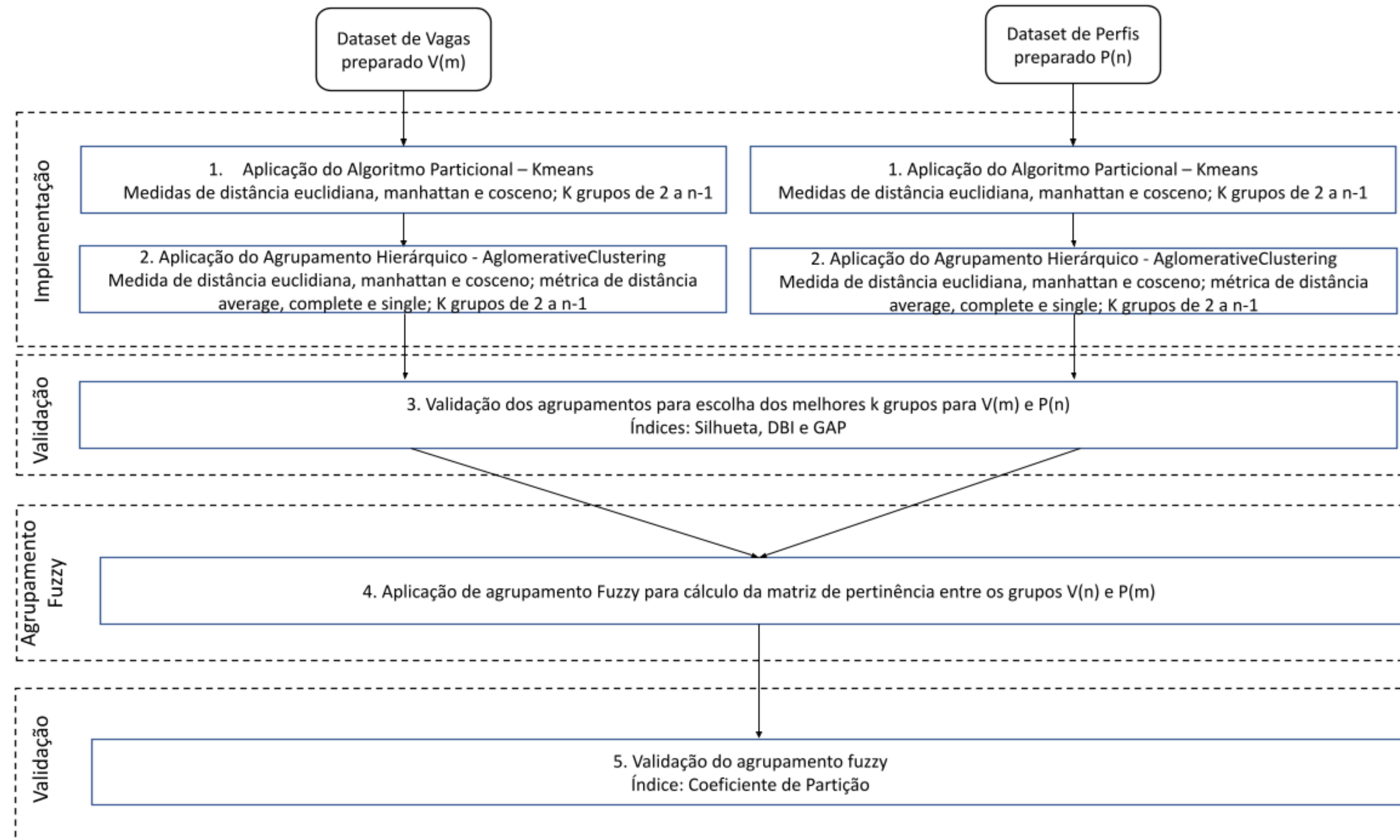
# Agenda

---



1. Contextualização do problema
2. Visão Geral da Proposta
3. Pré-processamento
4. Aplicação dos modelos de AM
5. Validação dos Resultados
6. Considerações e trabalhos futuros

# Aplicação dos modelos de AM



# Agenda

---



1. Contextualização do problema
2. Visão Geral da Proposta
3. Pré-processamento
4. Aplicação dos modelos de AM
5. Validação dos Resultados
6. Considerações e trabalhos futuros



# Validação dos Resultados



Objetivo	Script	Algoritmo / Agrupamento	Validação	Resultado em N° de Grupos
Limpeza dos dados	pre-processamento.ipynb	-	-	-
Agrupamento das Vagas	avaliacao-hiper-vagas.ipynb	Particional KMeans	Silhueta, DBI e GAP	9, 74, 74
		Hierárquico Aglomerativo	Silhueta, DBI	10, 74
Agrupamento dos Candidatos	avaliacao-hiper-candidatos.ipynb	Particional KMeans	Silhueta, DBI e GAP	10, 36, 49
		Hierárquico Aglomerativo	Silhueta e DBI	10, 74
Agrupamento de Vagas x Candidatos	avaliacao-hiper-fuzzy-c-means.ipynb	Fuzzy C Means	Coeficiente de Partição	14

- Scripts disponíveis em:
  - <https://github.com/abmorte/match-emprego>
  - <https://github.com/edmilsondejesus/match-emprego>

# Agenda

---



- 1. Contextualização do problema**
- 2. Visão Geral da Proposta**
- 3. Pré-processamento**
- 4. Aplicação dos modelos de AM**
- 5. Validação dos Resultados**
- 6. Considerações e trabalhos futuros**

# Considerações e trabalhos futuros



- A implementação demonstrou a possibilidade de outros cruzamentos e a classificação por graus de pertinência.
- Será necessário refinar os modelos e aprimorar índices de validação.
- Implementar outras formas de agrupamento/classificação.
- É viável a melhoria dos níveis de match através dos modelos de aprendizagem não supervisionada.

# Referências



- **EUSTÁQUIO, F. S. Um estudo sobre índices de validação de agrupamento fuzzy para dados de alta dimensionalidade. Universidade Federal da Bahia, 2017.**
- **IBGE. Desemprego. Disponível em: <<https://www.ibge.gov.br/explica/desemprego.php>>. Acesso em: 30 nov. 2022.**
- **J. C. B. Oliveira, R. A. Rios, E. S. de Almeida, C. N. Sant'Anna and T. N. Rios, Fuzzy Software Analyzer (FSA): A New Approach for Interpreting Source Code Versioning Repositories, 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2021, pp. 1-6, doi: 10.1109/FUZZ45933.2021.9494513**
- **PREVIDÊNCIA, M. DO T. E. Sine Aberto. Disponível em: <<https://sineaberto.economia.gov.br/>>. Acesso em: 30 nov. 2022.**
- **PREVIDÊNCIA, M. DO T. E. Portal Emprega Brasil. Disponível em: <<https://servicos.mte.gov.br/spme-v2/#/login>>. Acesso em: 30 nov. 2022.**



# Agradecimentos