

# SISCOD 3.0

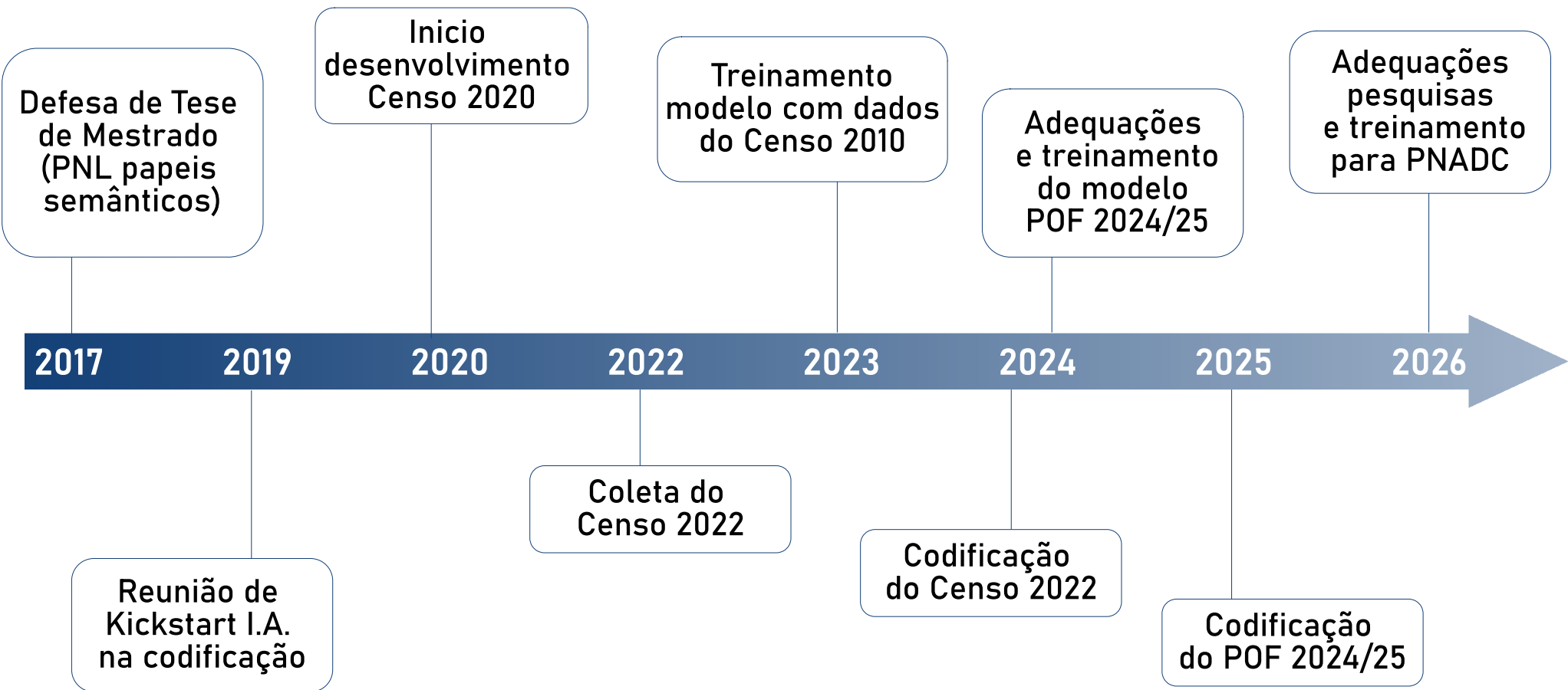
**Sistema de codificação de ocupações e atividades econômicas com I.A.**

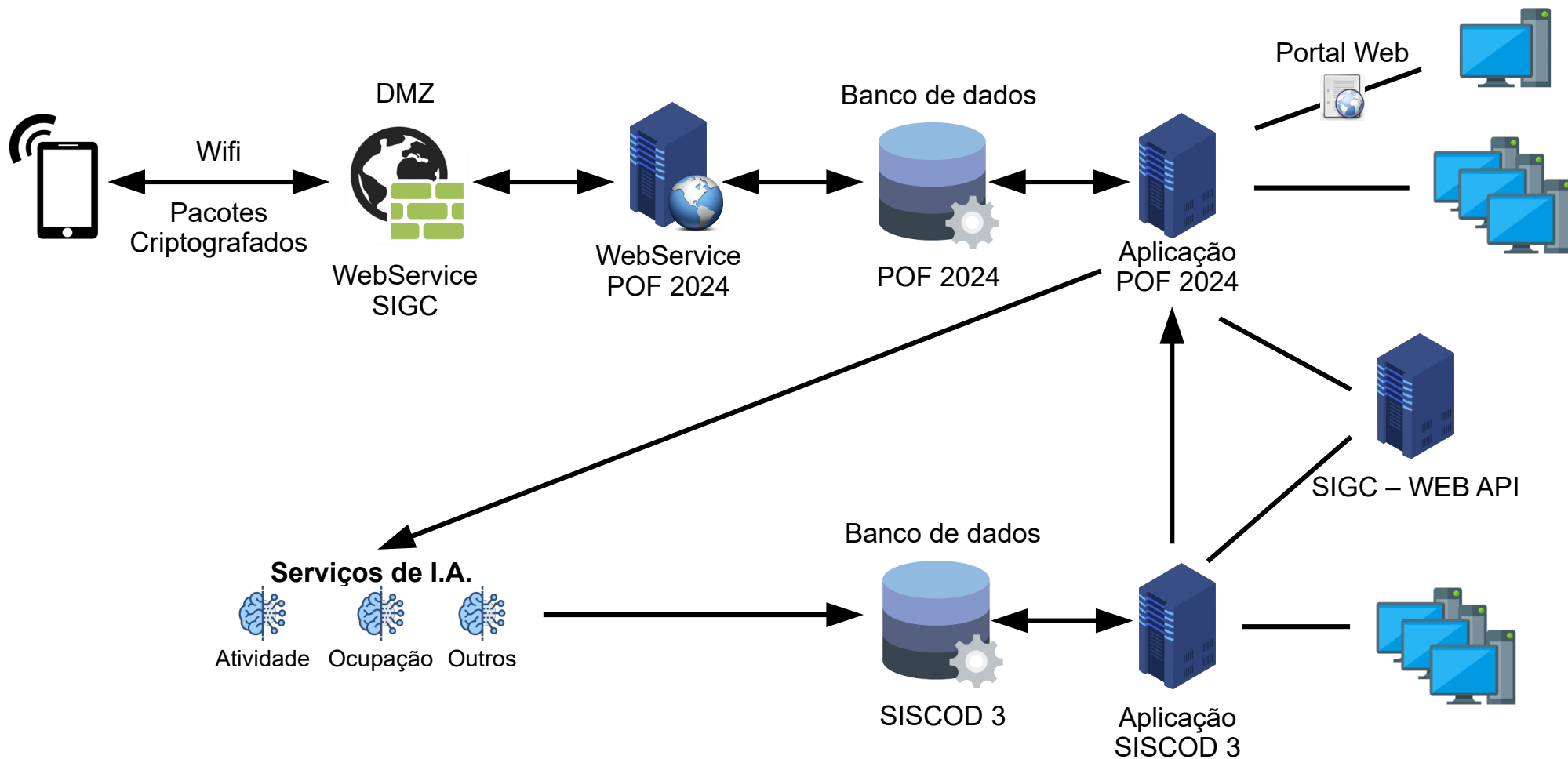
19/01/2026

Arthur Beltrão Castilho Neto



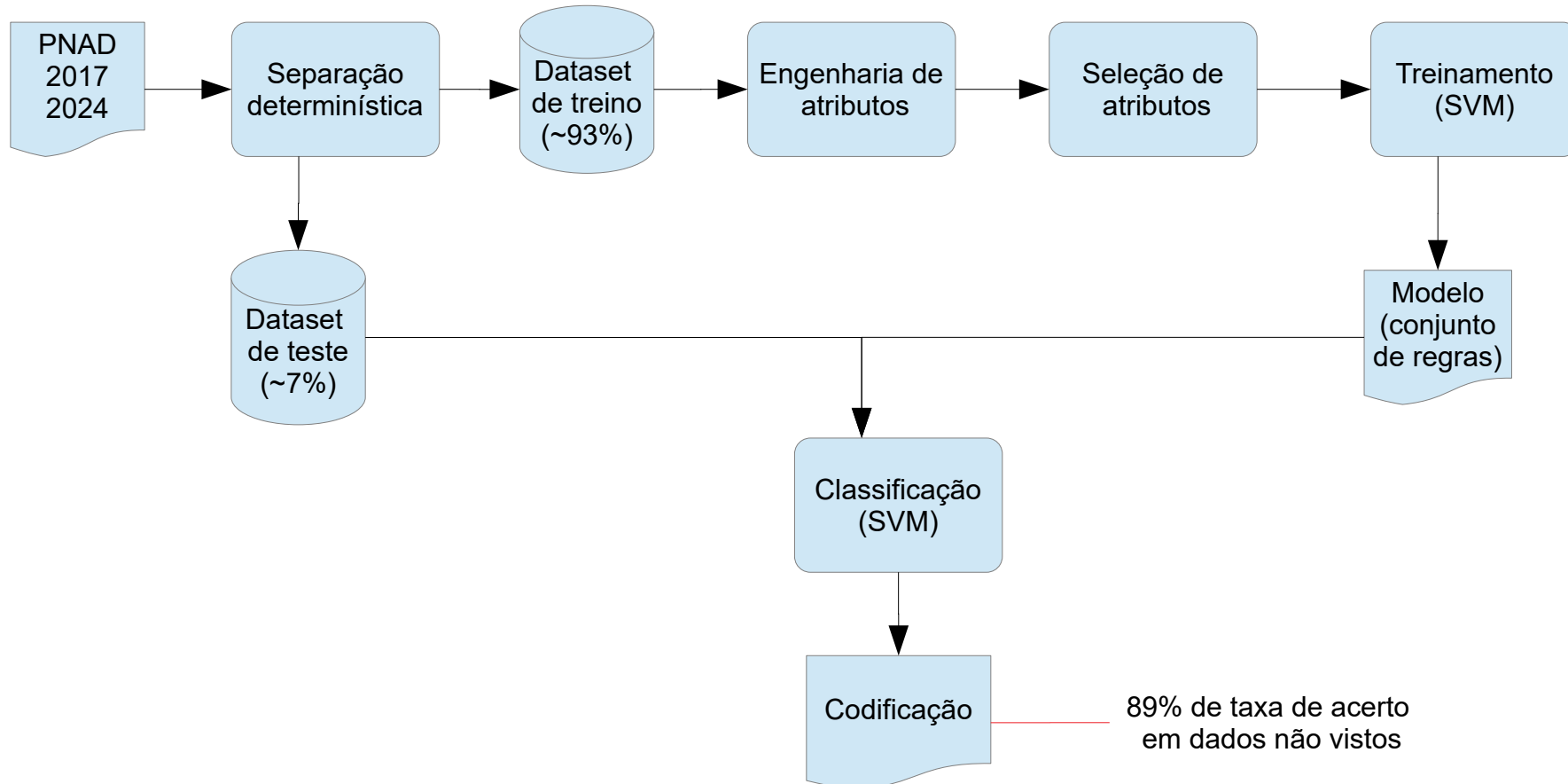
# Timeline do desenvolvimento





# Treinamento do Modelo

(Código Sugerido)

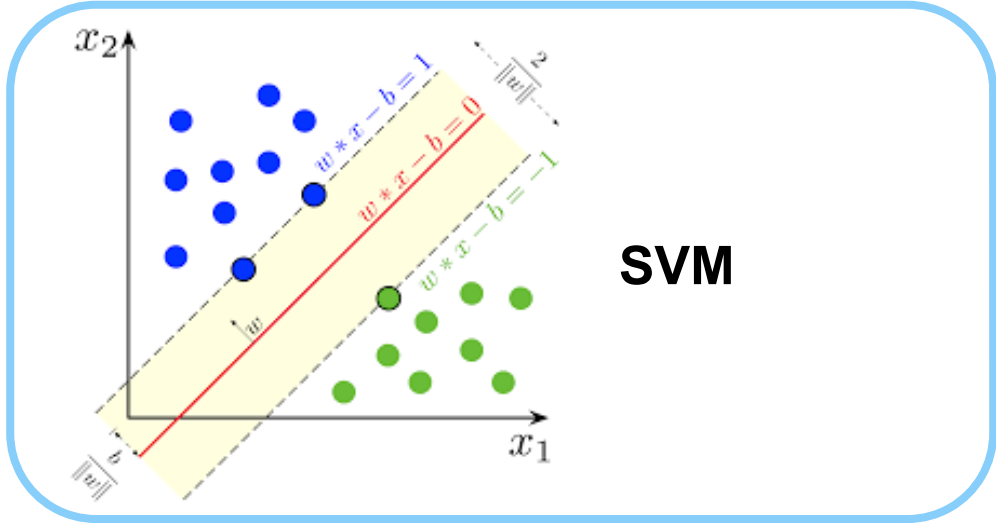


# Separação determinística

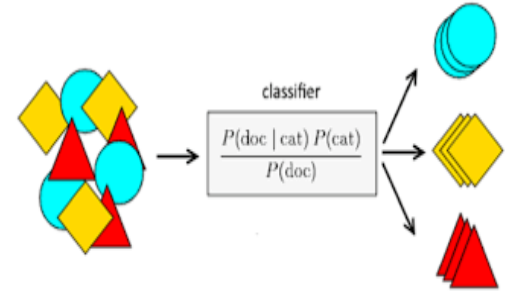
## Dados da PNAD Contínua



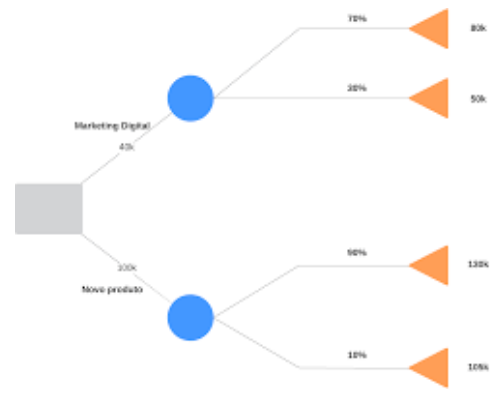
# Seleção do modelo



**SVM**



**Naive Bayes**



**Árvores de decisão**



**K-Means**

# Atributos

	the	red	dog	cat	eats	food
1. the red dog →	1	1	1	0	0	0
2. cat eats dog →	0	0	1	1	1	0
3. dog eats food →	0	0	1	0	1	1
4. red cat eats →	0	1	0	1	1	0

Bag of Words  
Ocupação e Atividade

$$1) \begin{cases} 3x - y = 7 \\ 5x + 2y = 8 \end{cases}$$

$$3x - y = 7$$

$$-y = 7 - 3x \cdot (-1)$$

$$y = -7 + 3x$$

$$5x + 2y = 8$$

$$5x + 2(-7 + 3x) = 8$$

$$5x - 14 + 6x = 8$$

$$11x = 8 + 14$$

$$11x = 22$$

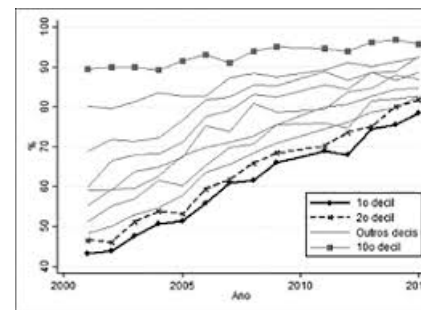
$$x = \frac{22}{11}$$

$$x = 2$$

$$y = -7 + 3 \cdot 2$$

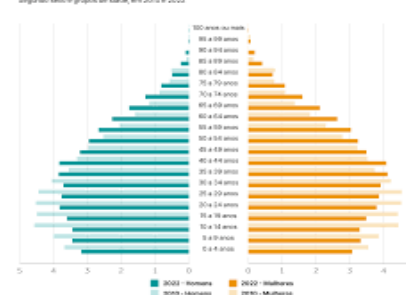
$$y = -7 + 6$$

Nível de Instrução



Faixa de rendimento

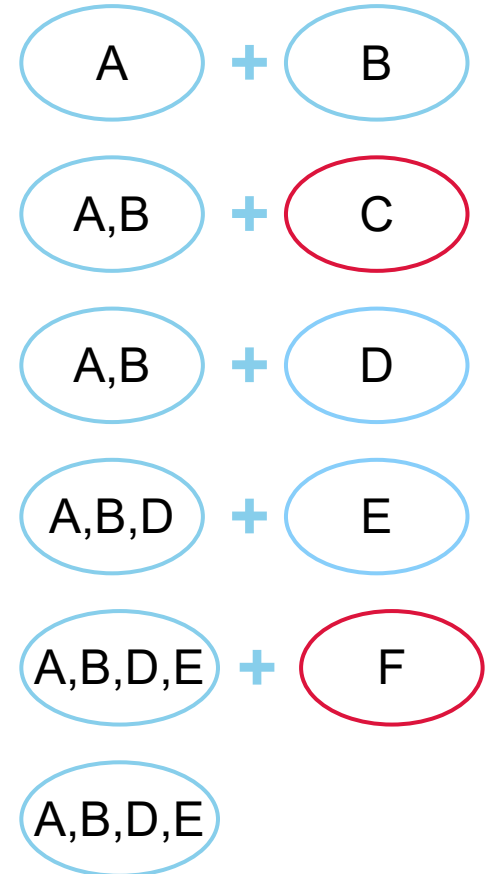
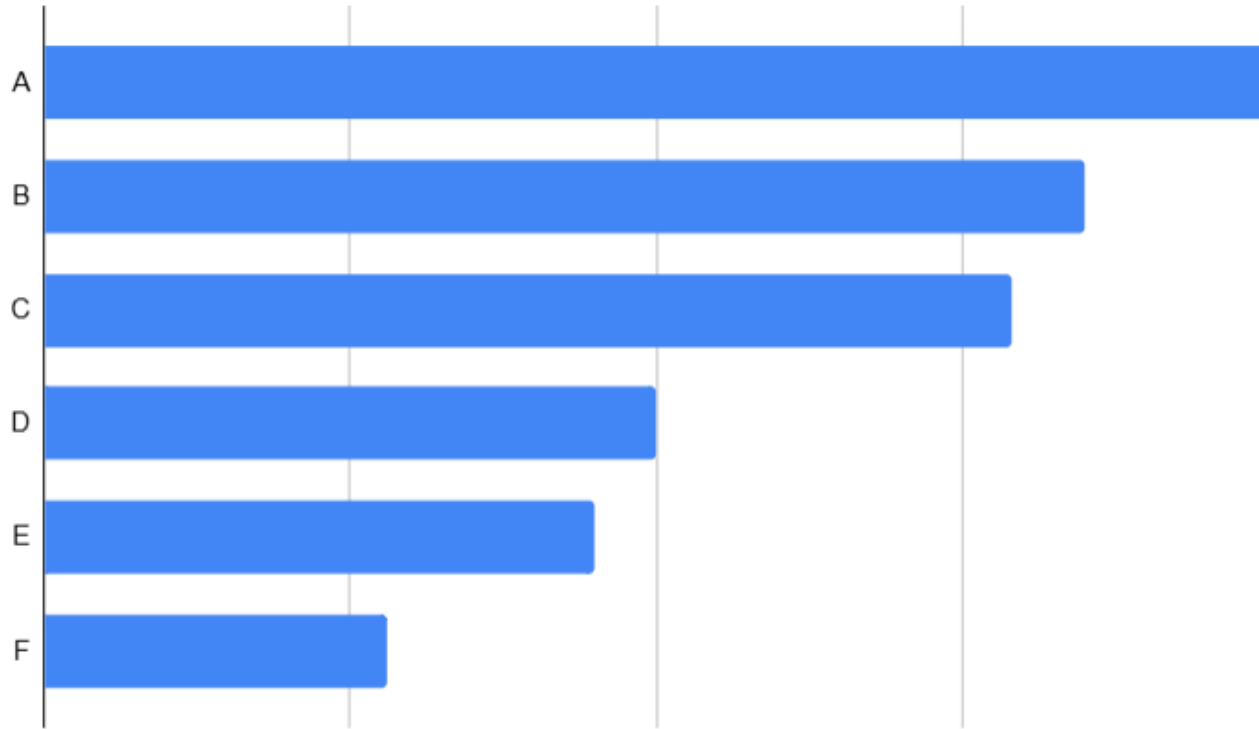
População residente no Brasil (%)



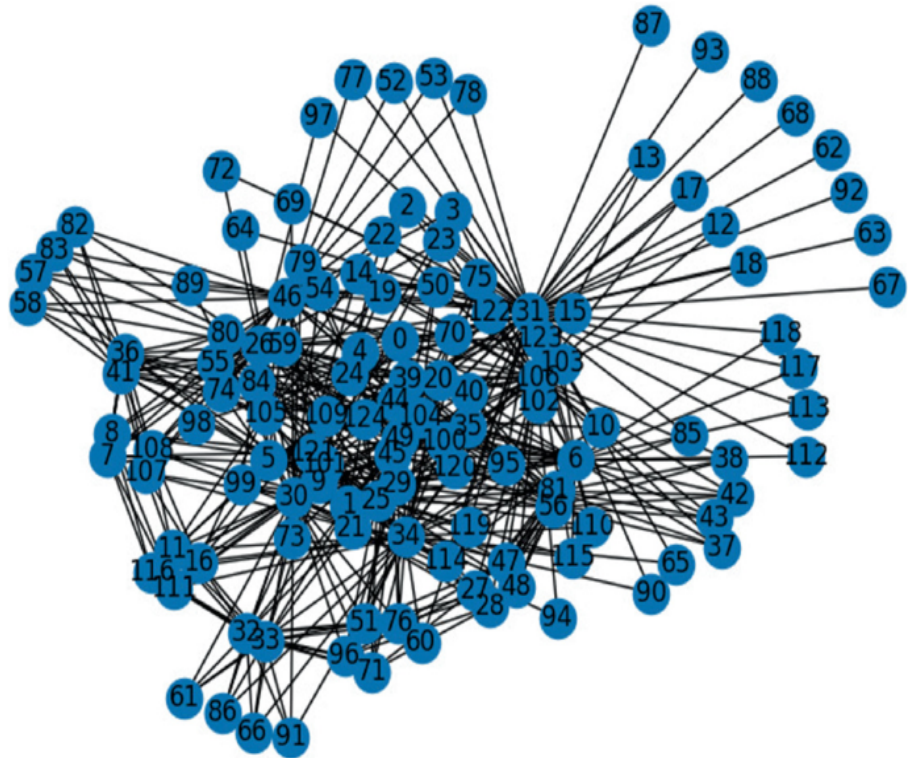
Fonte: Censo Demográfico 2010. População por idade e sexo. Resultados do universo. IBGE - Censo Demográfico 2010

Faixa etária

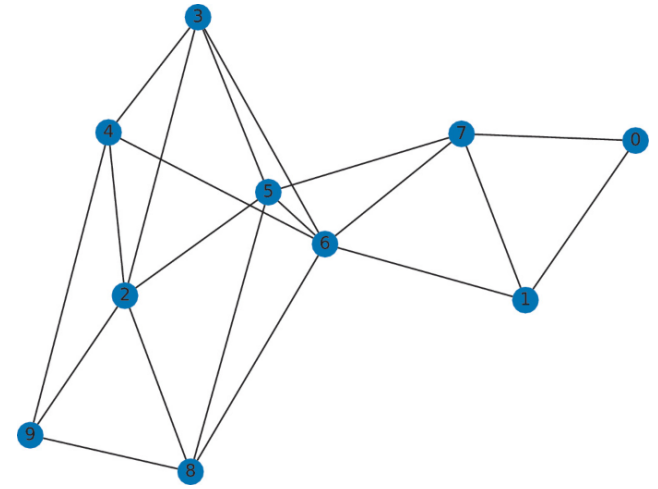
# Seleção: Greedy hill climbing



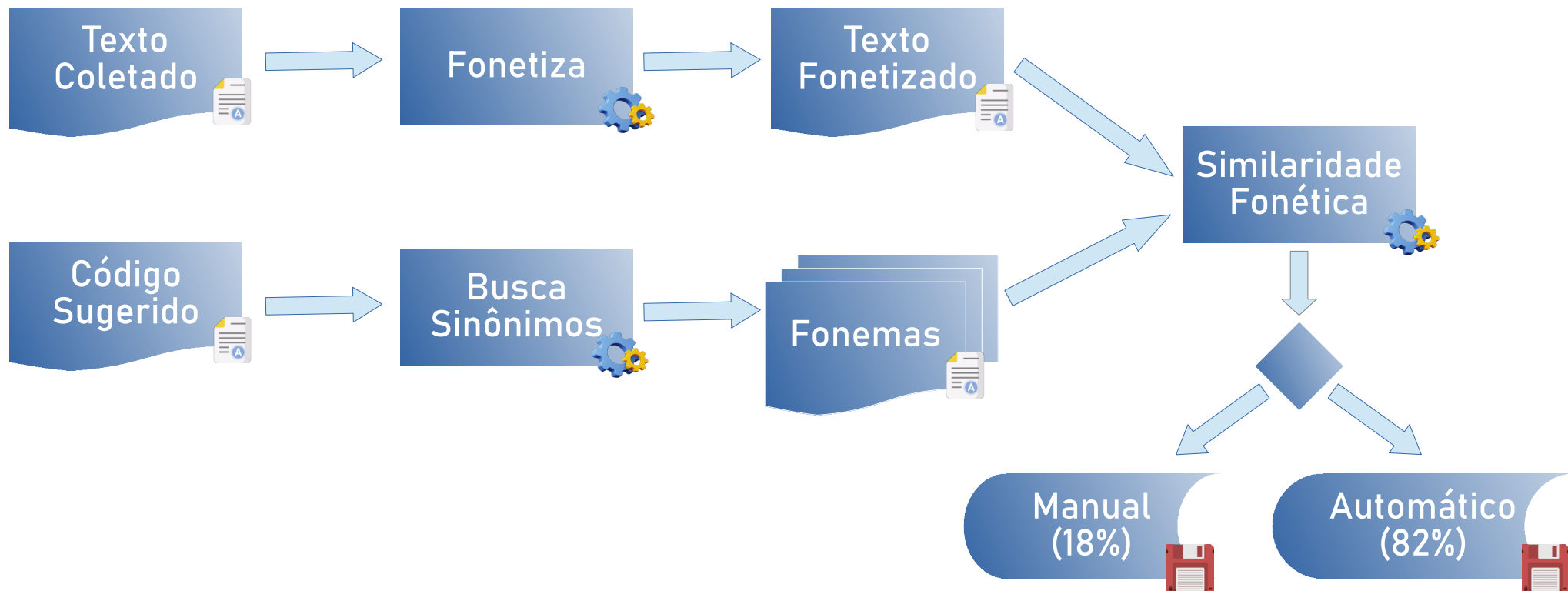
# Compactação do modelo SVM



89,10%



88,72%



## Ocupação

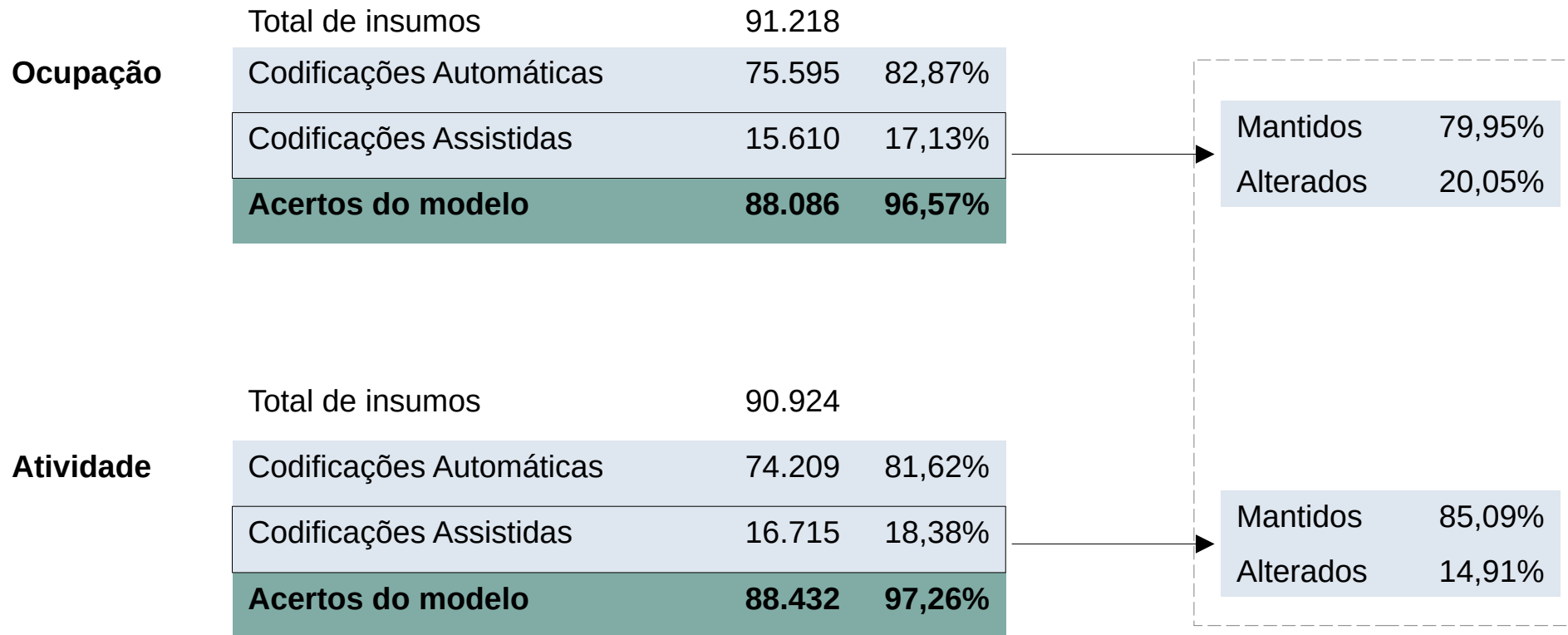
Insumos com DE/PARA	3.262.667	
Insumos sem DE/PARA	5.763.185	
Codificações Automáticas	4.278.472	74,24%
Codificações Assistidas	1.484.713	25,76%
<b>Acertos do modelo</b>	<b>5.128.322</b>	<b>88,98%</b>

Mantidos	57,24%
Alterados	42,76%

## Atividade

Insumos com DE/PARA	1.967.039	
Insumos sem DE/PARA	7.036.783	
Codificações Automáticas	3.944.158	56,05%
Codificações Assistidas	3.092.625	43,95%
<b>Acertos do modelo</b>	<b>5.848.596</b>	<b>83,11%</b>

Mantidos	61,58%
Alterados	38,42%



# Perguntas ?

