
Cofinanciado por:



UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

Projeto n.º **000193**

Apoio no âmbito do sistema de incentivos

Designação do projeto: **IA.: CIÊNCIA DOS DADOS NA CODIFICAÇÃO
AUTOMÁTICA e ANÁLISE PREDITIVA DE EXPRESSÕES**

Custo total elegível: 294 751,00 €

Apoio financeiro da UE: 167 860,69 € (56,95%)

Apoio financeiro público nacional: 126 890,30 € (43,05%)

Síntese do projeto:

O INE tem por Missão produzir, de forma independente e imparcial, informação estatística oficial de qualidade, relevante para a Sociedade, promovendo a coordenação, a análise, a inovação e a divulgação da atividade estatística nacional, garantindo o armazenamento integrado de dados.

O INE como uma Autoridade Estatística independente e credível, que desenvolve processos estatísticos metodologicamente avançados, que recorre à inovação tecnológica, à ciência de dados, à integração de múltiplas fontes para fins estatísticos, no respeito pela confidencialidade dos cidadãos e entidades, e que devolve à sociedade estatísticas de valor para um melhor conhecimento, investigação e a tomada de decisão.

O projeto tem como objetivo o tratamento e exploração de um grande volume de dados associado aos Censos2021 e aos inquéritos de periodicidade regular, através de técnicas avançadas de Inteligência Artificial e Ciência dos Dados, de forma a encontrar soluções para a

codificação automática de expressões bem como o autopreenchimento/análise preditiva das mesmas.

Considerando a situação atual, a operação visa atingir os seguintes objetivos:

- a) Reduzir significativamente a taxa de codificação manual das expressões contidas nas respostas das variáveis (de CAE, profissão e curso superior);
- b) Reduzir o tempo de tratamento de codificação da informação;
- c) Reduzir os custos associados à codificação manual de expressões;
- d) Transformação do processo de recolha de informação;
- e) Prestar novos serviços de recolha de informação aos cidadãos;
- f) Disponibilizar informação estatística de qualidade, de forma mais célere.

Dados a tratar:

Censos2021, Inquérito Piloto Censos 2020 e Inquéritos de periodicidade regular.

Técnicas a Usar:

Neste projeto pretende-se investigar e aplicar técnicas de análise e ciência dos dados nomeadamente:

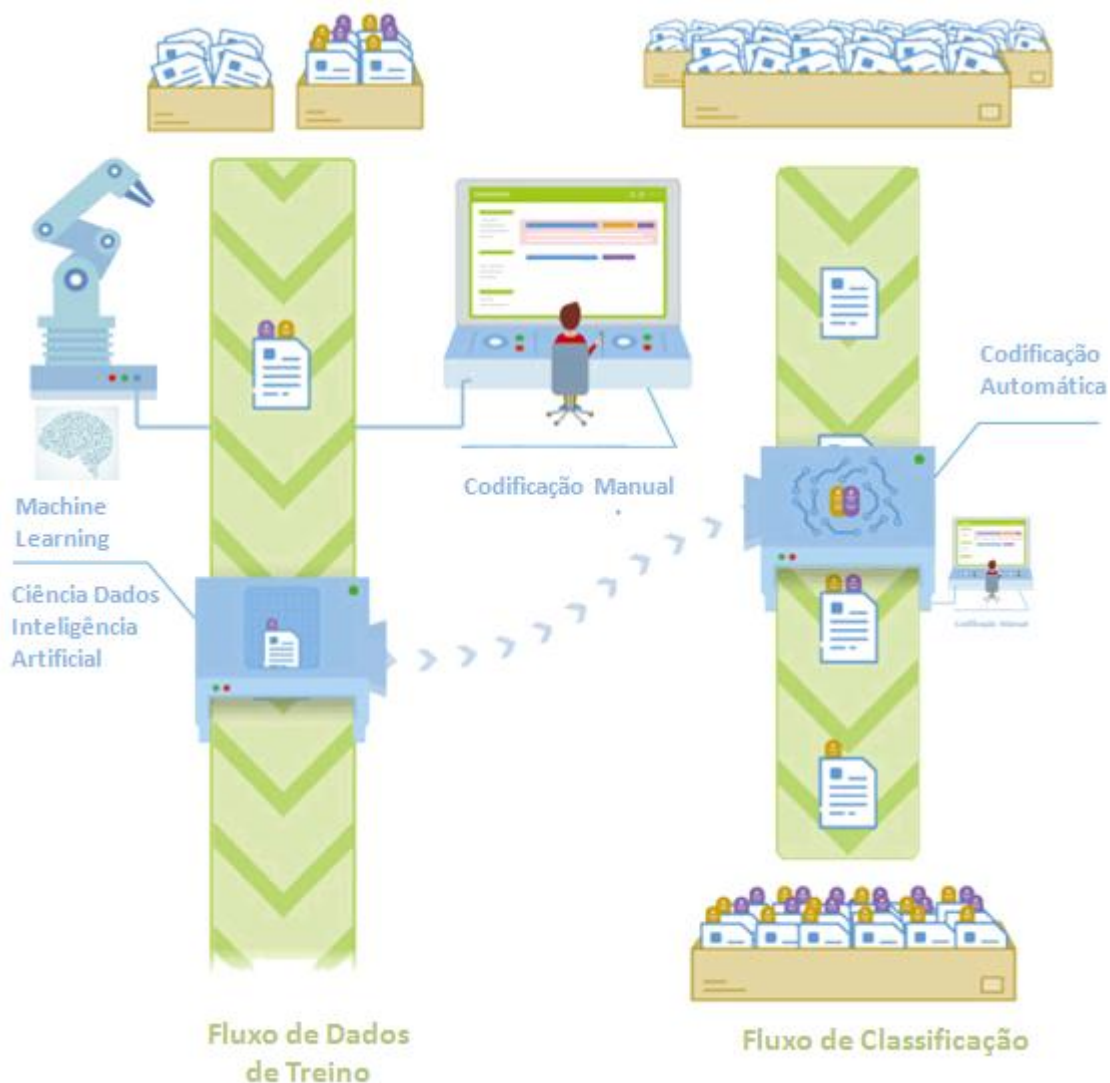
- Inteligência Artificial
- Ciência de Dados
- Machine Learning
- Extração de Conhecimento de Dados

Aplicados na :

- Pesquisa direta (“scrolling”)
- Árvores de pesquisa hierárquica
- Pesquisa por “matching” semântico ou simultâneo
- Pesquisa por palavras-chave
- Autopreenchimento/análise preditiva de expressões

Produtos e resultados esperados:

No final do projeto espera-se ser possível ter produzido algoritmos de decisão a aplicar a dados reais e modelos de autopreenchimento de expressões, que conduzam a uma efetiva redução das expressões a codificar manualmente, reduzindo os custos associados aos trabalhos manuais e uma redução no tempo de tratamento.



20-11-2019