



**Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Faculdade de Computação
Disciplina: Inteligência Artificial
Professor: Jefferson Moraes**

Trabalho Final

Agente Inteligente para Recomendação de Cargo na Área de Dados com Machine Learning e n8n

Objetivo

Neste trabalho, você irá construir um agente inteligente com apoio de aprendizado de máquina que sugere o cargo ideal na área de dados com base no perfil do usuário. O agente será implementado em um fluxo automatizado na plataforma n8n, integrando entrada de dados, pré-processamento e inferência com modelo treinado em Python ou outra linguagem de programação de sua escolha.

Instruções de Entrega

- Crie e entregue um único documento no formato Jupyter Notebook (.ipynb) usando a linguagem Python. Se utilizar outra linguagem de programação deve-se fornecer um documento descrevendo como a aplicação foi implementada e o código-fonte.
- O notebook deve conter: células de código com as implementações e células Markdown explicando cada etapa.
- O trabalho pode ser feito **individualmente ou em grupos de até 4 pessoas**.
- Prazo de entrega: **27 de agosto de 2025 até às 23:59h via SIGAA**.
- Apenas um integrante do grupo deve realizar a submissão.
- Se utilizar Google Colab: gere um link público, cole em um .txt e envie pelo SIGAA.
- Se feito localmente: compacte o .ipynb (ou o código-fonte de outra linguagem) e os dados em um .zip e envie.
- O notebook deve ser capaz de carregar automaticamente a base de dados fornecida..

Descrição do Dataset: State of Data Brazil 2022

A base de dados é proveniente do estudo State of Data Brazil 2022, conduzido pela Data Hackers e Bain & Company, com mais de 4200 respondentes. Ela inclui informações sobre perfil demográfico, formação, ferramentas utilizadas e cargos na área de dados.

O dataset original era composto por 4271 instâncias e 353 atributos. Após seleção e renomeação de alguns atributos, o dataset foi reduzido para 13 atributos listados na tabela abaixo:

#	Nome do atributo	Descrição	Tipo
1	idade	Idade do respondente	Inteiro
2	genero	Gênero do respondente	String
3	etnia	Etnia do respondente	String
4	pcd	O respondente é Pessoa com Deficiência?	String
5	vive_no_brasil	O respondente mora no Brasil?	Booleano
6	estado_moradia	Se mora no Brasil, qual o estado?	String
7	nivel_ensino	Nível de instrução do respondente	String
8	formacao	Formação acadêmica do respondente	String
9	tempo_experiencia_dados	Tempo de experiência do respondente na área de dados	String
10	linguagens_preferidas	Linguagens de programação preferidas do respondente	String
11	bancos_de_dados_preferidos	Ferramentas e sabores de bancos de dados preferidos do respondente	String
12	cloud_preferida	Ferramentas de cloud computing preferidas do respondente	String
13	cargo	Cargo atual ocupado pelo respondente	String

Você pode fazer o download do dataset neste link: <https://bit.ly/facomp-sods>.

Utilize a base de dados fornecida e aplique o processo KDD para realizar uma tarefa de classificação de dados, sendo livre o uso de bibliotecas e frameworks. A classificação consiste no uso dos 12 primeiros atributos (de idade até cloud_preferida) para encontrar qual o cargo ideal para um candidato a uma vaga na área de dados. As tarefas de cada etapa do processo KDD a serem cumpridas estão descritas a seguir.

Etapas do Projeto

1. Seleção e Análise Exploratória

- Justifique a escolha ou manutenção dos atributos mais relevantes.
- Se usar todos os dados, justifique.
- Faça uma análise exploratória com visualizações que mostrem:
 - Distribuições
 - Correlações
 - Valores ausentes
 - Outliers ou ruídos

2. Pré-processamento e Enriquecimento

- Trate dados ausentes, inconsistentes ou ruidosos.
- Crie atributos derivados, se pertinente (ex: transformar tempo de experiência em número).
- Normalize ou codifique atributos categóricos conforme a técnica de ML escolhida.

3. Formatação e Divisão de Dados

- Aplique divisão de dados hold-out estratificada (70% treino / 30% teste), mantendo a proporção das classes da variável cargo.

4. Mineração de Dados (Machine Learning)

- Escolha e justifique duas técnicas entre as abaixo:
 - Regressão Logística
 - Árvore de Decisão
 - Floresta Aleatória
 - Support Vector Machine (SVM)
 - Rede Neural
 -
- Treine os modelos e avalie com:
 - Acurácia
 - Precisão
 - Recall
 - Matriz de confusão
- Justifique os hiperparâmetros utilizados.

5. Construção de Agente Inteligente no n8n

Implemente no n8n um agente que:

- Receba inputs de um usuário (via Webhook, Google Forms ou Telegram Bot);
- Realize o pré-processamento do input;
- Carregue o modelo treinado (via HTTP Request, webhook local ou API externa);
- Execute a classificação do cargo ideal com base nos dados fornecidos;
- Retorne a recomendação ao usuário (via Telegram, e-mail ou dashboard).

Você pode usar o HTTP Request para acionar uma API com o modelo, ou integrar com um script Python hospedado no Hugging Face Spaces, Replit ou em backend próprio.

6. Documentação da Solução

O notebook ou outro documento deve incluir:

- Explicações das decisões tomadas;
- Código do modelo e visualizações;
- Screenshots do fluxo no n8n em funcionamento;
- Link para teste (se aplicável);
- O .json exportado do fluxo do n8n.

Avaliação

Critério	Pontos
Modelagem e avaliação dos algoritmos de IA	4,0 pontos
Automação e integração via n8n	3,0 pontos
Clareza e qualidade do notebook	2,0 pontos
Criatividade e inovação (extras, front-end etc.)	1,0 ponto
Total	10 pontos