

🤖 Projeto: Inteligência Artificial para Recrutamento — Decision AI (Datathon FIAP)

📌 Visão Geral

Este projeto foi desenvolvido para o **Datathon Decision - Pós Tech FIAP**, com o objetivo de construir uma **solução de Inteligência Artificial aplicada ao Recrutamento e Seleção** de profissionais de TI.

A aplicação integra técnicas de **Machine Learning supervisionado**, **similaridade semântica com embeddings**, e uma API robusta desenvolvida com **FastAPI**, permitindo que a empresa **Decision** otimize o processo de triagem, entrevistas e contratação de talentos.

🧩 Problemas Enfrentados pela Decision

- Falta de padronização nas entrevistas
- Dificuldade em medir aderência técnica dos candidatos
- Longo tempo de triagem de perfis
- Match impreciso entre candidatos e vagas

🎯 Objetivo da Solução

- Automatizar a triagem de currículos com alto grau de acurácia
- Gerar um ranking de compatibilidade técnica entre candidatos e vagas
- Reduzir o tempo do processo seletivo
- Apoiar a tomada de decisão com base em dados e métricas

🛠️ Arquitetura da Solução

🔄 Pipeline Inteligente de Recrutamento

1. **Data Collection**

Leitura de grandes volumes de dados históricos (` applicants.json` , ` prospects.json` , ` vagas.json`).

2. **Pré-processamento e Construção do Dataset Supervisionado**

- Padronização de campos
- Tratamento de valores nulos
- Criação de dataset para classificação binária (` match = 1 ou 0`)

3. **Engenharia de Features**

- TF-IDF sobre currículos
- Codificação de variáveis categóricas
- Vetorização de textos com embeddings SBERT (semantic match)

4. **Modelo Supervisionado**

- **XGBoost Classifier**
- Aplicação de **SMOTE**
- Ajuste de ` scale_pos_weight` e ` threshold`

5. **Geração de Relatório**

- Geração de relatório em HTML com métricas como: F1-score, recall da classe positiva, curva ROC

6. **API com FastAPI**

- Endpoints para classificação e ranking técnico
- Respostas com score e status de compatibilidade

7. **Interface com Streamlit**

- Interface amigável para RH testar candidatos e visualizar rankings

🗖️ Tecnologias Utilizadas

Categoria	Ferramentas	
-----	-----	
Backend/API	FastAPI, Uvicorn	
Machine Learning	XGBoost, Scikit-learn, imbalanced-learn	
Similaridade Semântica	SentenceTransformers (SBERT)	
Deploy/Container	Docker, GitHub Actions	
Logging & Relatórios	logging, HTML + Jinja2	
Frontend	Streamlit	
Testes	Pytest	

🌐 Endpoints da API

Método	Rota	Descrição	
-----	-----	-----	
POST	`/match`	Recebe dados de um candidato e retorna a compatibilidade técnica semântica	
POST	`/rank`	Recebe lista de candidatos e uma vaga; retorna ranking por compatibilidade	

💡 Exemplo de Payload para `/match`

```
```json
{
```

```

"nome": "Carlos Mendes",

"cv": "Consultor SAP BASIS com experiência em ambientes AWS e Oracle. Responsável
por liderar suporte técnico e implantações. Inglês fluente.",

"nivel_ingles": "Fluente",

"area_atuacao": "TI - Sistemas e Ferramentas-"
}
` `` `

```

### ♦ Exemplo de Payload para `/rank`

```

` `` `json
{
 "id_vaga": "5185",
 "candidatos": [
 {
 "nome": "Carlos Mendes",

 "cv": "Consultor SAP BASIS com experiência em ambientes AWS e Oracle. Responsável
por liderar suporte técnico e implantações. Inglês fluente.",

 "nivel_ingles": "Fluente",

 "area_atuacao": "TI - Sistemas e Ferramentas-"
 },
 {
 "nome": "Rodrigo Lima",

 "cv": "Especialista em operações de infraestrutura com foco em gestão de
fornecedores e controle de SLAs. Experiência com SQL e gestão de custos.",

 "nivel_ingles": "Avançado",

 "area_atuacao": "TI - Sistemas e Ferramentas-"
 },
 {
 "nome": "João da Silva",

 "cv": "Profissional com ampla experiência em gestão de operações de TI. Liderou
equipes em projetos de Cloud, com expertise em AWS, SAP BASIS, banco de dados SQL e

```

Oracle. Fluente em inglês. Forte habilidade em gerenciamento de SLA e relacionamento com clientes.",

```
"nivel_ingles": "Avançado",
"area_atuacao": "TI - Sistemas e Ferramentas-"
},
{
 "nome": "Maria Oliveira",
 "cv": "Engenheira de software com foco em Java e sistemas bancários. Experiência em liderança técnica e projetos internacionais. Espanhol avançado.",
 "nivel_ingles": "Básico",
 "area_atuacao": "TI - Desenvolvimento de Software"
},
{
 "nome": "Ana Beatriz",
 "cv": "Analista de RH com ênfase em recrutamento tech. Conhecimento básico em sistemas de ERP.",
 "nivel_ingles": "Intermediário",
 "area_atuacao": "Recursos Humanos"
}
]
}
...

```

## ## 🧪 Testes Automatizados

Localizados em `backend/tests/` e `frontend/tests/`, cobrindo:

- \* Pipeline de dados
- \* Feature engineering
- \* Treinamento de modelo

\* API (match e rank)

\* Interface com Streamlit

---

## ## 📊 Relatórios de Métricas

Gerados automaticamente via ``metric_report.py`` :

\* Salvos em: ``backend/metric_reports/``

\* Incluem: curva ROC, classificação por threshold, recall da classe 1, etc.

---

## ## 📁 Estrutura do Projeto

```

fiap-posmlps-fase5-datathon-decision

```
├── .github/
│   ├── workflows/
│   └── pipeline.yaml
│
├── backend/
│   ├── app/
│   │   ├── etl/
│   │   │   ├── backblaze_loader.py
│   │   │   ├── data_collector.py
│   │   │   └── data_preprocessing.py
│   │   └──
│   └── model/
│       └── feature_engineering.py
```

```

| | | ├── model_evaluation.py
| | | ├── model_training.py
| | |
| | ├── report/
| | | ├── metric_report.py
| | |
| | ├── semantic/
| | | ├── semantic_api_matcher.py
| | | ├── semantic_dataset_builder.py
| | | ├── semantic_matcher.py
| | |
| | ├── utils/
| | | ├── constants.py
| | | ├── model_utils.py
| | |
| | ├── api.py
| |
| ├── data/
| | ├── dataset_processado.csv
| | └── features_treinamento.pkl
| |
| ├── logs/
| | └── app.log
| |
| ├── metric_reports/
| | └── model_metric_report_<timestamp>.html
| |
| ├── models/
| | ├── model.pkl
| | └── feature_pipeline.pkl

```

Frontend (Streamlit)


```
```bash
cd frontend
docker build -t decision-frontend .
docker run -p 8501:8501 decision-frontend
```
```

Interface: http://localhost:8501

Entregáveis do Projeto

- ✓ Modelo supervisionado treinado e testado
- ✓ API REST funcional com FastAPI
- ✓ Sistema semântico de match com embeddings
- ✓ Relatórios HTML com métricas completas
- ✓ Logging centralizado e estruturado
- ✓ Frontend com Streamlit para teste de perfis
- ✓ Testes automatizados para API e pipeline
- ✓ Pronto para deploy via Docker e Railway

Vídeo Demonstrativo

 \[Inserir link para o vídeo explicativo da solução]

 **Autoria**

Projeto desenvolvido por:

****Alexandro de Paula Barros****

Para o ****Datathon Pós Tech FIAP – Decision****