



Mariana Duarte

CIENTISTA DE DADOS

Lisboa, Portugal
mduarte14@gmail.com
+351 912 538 647
linkedin.com/in/marianaduarte-ds

Cientista de dados com cinco anos de experiencia no desenvolvimento de modelos de machine learning para problemas de negocio em larga escala. Na Talkdesk, construi modelos de NLP que classificam automaticamente 14 milhoes de interacoes mensais de suporte. Formacao solida em estatistica, Python e MLOps, com publicacoes em conferencias da especialidade.

EDUCAÇÃO

Mestrado in Ciencia de Dados e Engenharia

Instituto Superior Tecnico, Lisboa

Set. de 2017 – Jul. de 2019

Dissertacao sobre *modelos de deep learning para analise de sentimento em portugues europeu*. Media de 18 valores.

Licenciatura in Matematica Aplicada e Computacao

Instituto Superior Tecnico, Lisboa

Set. de 2014 – Jul. de 2017

Media final de 17 valores. Premio de melhor aluna do curso em 2016/2017.

HABILIDADES

- Python (scikit-learn, TensorFlow, PyTorch)
- SQL & PySpark
- NLP (Hugging Face, spaCy)
- MLflow & Kubeflow
- AWS (SageMaker, EC2, S3)
- Docker & Kubernetes
- Estatistica e Probabilidade
- Power BI & Matplotlib
- Git & CI/CD
- Pandas & NumPy

CERTIFICAÇÕES

AWS Certified Machine Learning – Specialty

Amazon Web Services

Ago. de 2023 – Ago. de 2026

TensorFlow Developer Certificate

Google

Mai. de 2022 – Mai. de 2025

IDIOMAS

- Portugues (nativo)
- Ingles (fluyente – C2)
- Frances (intermedio – B1)

REFERÊNCIAS

Ricardo Guerreiro

Head of AI, Talkdesk

r.guerreiro@talkdesk.com, +351 918 432 761

Helena Matos

Senior Research Scientist, Feedzai

h.matos@feedzai.com, +351 927 653 418

EXPERIÊNCIA

Cientista de Dados, Talkdesk, Lisboa

Jan. de 2022 – Presente

Equipa de AI e machine learning do produto de contact center.

- Desenvolvi modelo de classificacao de intencoes em NLP que processa **14 milhoes de interacoes/mes** com accuracy de 91%
- Implementei pipeline de MLOps com MLflow e Kubeflow para **8 modelos em producao**
- Criei modelo de previsao de churn que identificou **73% dos cancelamentos** com 30 dias de antecedencia
- Reduzi tempo de treino de modelos em **58%** migrando para GPUs na AWS

Cientista de Dados Junior, Feedzai, Coimbra

Mar. de 2020 – Dez. de 2021

Desenvolvimento de modelos de detecao de fraude para o setor financeiro.

- Contribui para modelo de detecao de fraude em tempo real com **precision de 96%** e recall de 89%
- Processei datasets de **2,3 mil milhoes de transacoes** para feature engineering
- Desenvolvi pipeline de dados em PySpark para preparacao de features em **batch e streaming**

Investigadora de Machine Learning, INESC-ID Lisboa, Lisboa

Jan. de 2019 – Fev. de 2020

Investigacao em processamento de linguagem natural para a lingua portuguesa.

- Co-autora de **2 artigos** publicados em conferencias internacionais (ACL, EMNLP)
- Desenvolvi corpus anotado de sentimento em portugues europeu com **45.000 frases**

PROJETOS

Classificador de Intencoes NLP – Talkdesk

Jun. de 2022 – Abr. de 2023

Modelo de classificacao de intencoes para contact centers multilingue.

- Accuracy de **91%** em 6 linguas, processando 14 milhoes de interacoes/mes
- Modelo baseado em fine-tuning de **BERT multilingue** com dados proprietarios
- Reducao de **34% no tempo medio** de routing de chamadas

Corpus de Sentimento PT-EU – INESC-ID

Mar. de 2019 – Jan. de 2020

Criacao de corpus anotado de sentimento para portugues europeu.

- Dataset de **45.000 frases** anotadas com sentimento e emocao
- Publicado e disponibilizado para a comunidade academica
- Citado em **12 artigos** internacionais

ATIVIDADE EXTRACURRICULAR

Oradora – Data Science Portugal Meetup

Jun. de 2022

Apresentacoes sobre NLP e MLOps em meetups trimestrais. Audiencia media de 90 participantes.