António Rocha

Data Scientist

✉ workbrigade@gmail.com

📞 +351 963286251

📍 Lisbon, Portugal

in António Rocha

Profile

Data scientist with 2-3 years of experience, specializing in natural language processing (NLP). I enjoy fostering a pleasant collaborative work environment and actively seek to learn from colleagues and feedback. With decent skills in machine learning (ML), I possess the knowledge needed to build ML products end-to-end.

Interests

Dance | Crossfit | Read

Languages

Portuguese
Native
● ● ● ● ●

English
● ● ● ● ●

French
Currently taking classes
● ● ○ ○ ○

Education

Bachelor's Degree in computer science and management
ULHT - Lusophone University of Humanities and Technologies ⤢
present | Lisbon, Portugal
This degree aims to develop IT profiles able to interconnect businesses with companies, such as:
- IT consultant
- Business support analyst
- Project Manager
- IS Architect
- Enterprise Application Development

Professional Training in Web Development
ISCTE – University Institute of Lisbon
2020 – 2021 | Lisbon, Portugal
Professional Training, with a duration of 800 hours. The training included a theoretical-practical component and on-the-job training lasting 3 months.
- Programming with Python
- Web Programming (html/css/Javascript)
- Mathematics for Python
- Data base
- Web server programming
- Web services development
- Job market skills

Professional Experience

Data Scientist
Axians ⤢
May 2021 – Jul 2023 | Lisbon, Portugal

Full time professional programmer working on:

- Exploratory Data Analysis (EDA).
- Research and development of machine learning and deep learning models.
- Training and testing of models in Python.
- Unit and integration testing.
- Development of REST services (APIs).
- Presentation of results and findings to clients.
- Deployment of the models in productive environments.
- Any necessary work directly related to the points mentioned above.

Programing Languages

Python   ● ● ● ● ○

SQL   ● ● ● ○ ○

Bash/Shell   ● ● ● ○ ○

Html   ● ● ● ● ○

CSS   ● ● ○ ○ ○

Java   ● ● ○ ○ ○

Javascript   ● ● ○ ○ ○

C   ● ● ○ ○ ○

R   ● ● ○ ○ ○

Skills

Requirement analysis

Software Development

Agile Methods

Testing

Web Development

Microservices

Databases management

Web Scrapping

Math

Statistics

A/B Testing

Data Analysis

Data Visualization

Research

Machine Learning

Deep Learning

MLOps

UX/UI

Accounting

Finances

Marketing

Projects

CV Parser
Extraction with unknown layouts
2022 – present
Currently working in an ongoing self-managed project within the academic domain (college) to develop a CV parser. Leveraging Faster R-CNN for object detection and Named Entity Recognition (NER).

Main highlights of the project:

- Explored a variety of tools for the Data Science workflow.
- Explored the field of computer vision for machine learning and deep learning.
- Conducted web scraping to collect CVs from the internet as the project dataset.
- Utilized the Faster R-CNN model to predict bounding boxes for different CV elements.
- Extracted relevant information by applying NER techniques to the text within the bounding boxes.
- Continuously working to improve the CV parser's performance and adaptability to handle diverse CV layouts and languages.

Social Security V360
FAQs Search Engine in Portuguese
2022 – 2023

Developed a search engine utilizing machine learning techniques from a collection of Frequently Asked Questions (FAQs) written in Portuguese.

Main highlights of the project:

- Self-managed project lifecycle, collaborating with stakeholders.
- Gathered, augmented, cleaned and preprocessed a dataset of Portuguese FAQs.
- Utilized vector databases, applied statistical techniques and word embeddings to create numerical representations of FAQs.
- Developed a machine learning model for information retrieval with NLP.
- Implemented query preprocessing, ranking, and retrieval of relevant FAQs.
- Reduced model size using ONNX format and quantization techniques.
- Deployed the model using Docker, Kubernetes and Rancher for easy scalability and high availability across environments.
- Helped implement a user-friendly interface for query input, used in client demos.
- Evaluated the system's performance using metrics like accuracy, precision, recall, mean reciprocal rank, and others.

Technology Stack (Libraries, Frameworks and Tools)

Data Storage
PostgreSQL, MinIO, Feast

Experiment Tracking
MLFlow

Test and Automation
Jenkins, Pytest, Selenium

Observability
Prometheus, Grafana, ElasticSearch

Machine Learning
Scikit-Learn, XGBoost, OpenCV

Deep Learning
PyTorch, TensorFlow, Keras

Data Manipulation
Pandas, Dask

Statistical Modeling
Statsmodels, SciPy

Numerical Computation
Numpy

Data Visualization
Matplotlib, Seaborn

Web Development
Django, FastAPI, Flask

Interoperability
ONNX

Office Productivity
Microsoft Office (Excel, PowerPoint, Word)

Cloud Platforms
Azure, AWS

DevOps
Git, Airflow, Docker, Kubernetes

- Implemented some feedback loop mechanisms for performance enhancement, analysing search logs, and monitoring computer resources.

Tax Authority eKnowledge Assistant
CNN Classifier for text
2021 – 2022

Developed a Convolutional Neural Network (CNN) model for interpretation and classification of tax-related requests.

Main highlights of the project:

- Conducted research on various models.
- Built a CNN model, leveraging deep learning libraries.
- Integrated the CNN model through APIs into the existing application architecture.
- Actively helped design and execute tests to ensure the reliability and accuracy of the model.
- Contributed to deploying the application on Azure, setting up infrastructure and configuring resources for optimal performance and availability.

Soccer Match Prediction
Sports Betting Profitability
2018
Developed a predictive model to increase the profitability of betting on soccer matches.

Main highlights of the project:

- Used regression modelling to develop a predictive model.
- Integrated Poisson distribution to predict expected goals.
- Performed an evaluation using statistical techniques.
- Performed profitability analysis by monitoring and evaluating real betting results and adjusted the model to increase profitability.
- Implemented using Microsoft Excel.

Certificates

- Data Scientist with Python (Career Track Datacamp) ⬀

- Machine Learning Scientist (Career Track Datacamp) ⬀

- Data Engineer (Career Track Datacamp) ⬀

Awards

2º Place Breakthrough Innovation Challenge 2022
Axians