VITOR SOUSA - Data Scientist

vitorhcsousa | vitorhcsousa | 917123945 | vitorsousavhcs@outlook.com | vitorhcsousa.github.io

EXPERIENCE

Data Scientist FARFETCH — Jun 2022 – Present — Porto, Portugal

- Developed advanced size recommendation models utilizing LSTM, GRUs, and attention mechanisms, resulting in a significant enhancement of top1-accuracy by **22%-45%** compared to previous solutions.
- Spearheaded a solo Proof of Concept (POC) project to engineer an automated Size & Fit recommendation model. Through manual verification, the model has been validated and is now operational, markedly improving product recommendations and user experience across a portfolio of over 25,000 items.
  - The model yielded tangible results, notably **reducing return rates by 7% in Men's apparel and 3.6% in Women's apparel**.
- Collaborated in the design of machine learning models addressing critical business challenges within the e-commerce landscape.
  - Conducted various ablations, such as integrating novel signals like "add to bag," resulting in a **24.5%** increase in user coverage.
- Developed and managed ETL pipelines in production, ensuring data integrity and reliability.
- Prototyped, evaluated, and deployed models into production environments, enhancing operational efficiency and business processes.
- Designed and conducted AB tests and Causal Impact analyses to assess the efficacy of new features and algorithms.
- Contributed to the conceptualization and strategic planning of future projects, aligning with organizational goals and priorities.

Python | SQL | PySpark | Databricks | Azure | Airflow | Terraform | Git | MlFlow | Looker | BigQuery | PyTorch | LSTM | Neural Networks | Docker | Bayesian Models

Data Scientist Sonae MC — May 2021 – Jun 2022 — Porto, Portugal

- Engineered predictive models (RFM, behaviour segmentation, churn prediction, sentiment analysis) using diverse data sources, notably Continente Card, Portugal's premier loyalty program.
- Enhanced marketing impact predictions and strategies by blending machine learning expertise with business acumen.
- Developed an interactive Streamlit platform for analyzing models, empowering cross-functional business teams.
- Successfully migrated models from SAS to Python/PySpark on Azure, ensuring robust deployment and ongoing maintenance.

Python | SAS | SQL | PySpark | Databricks | Azure | Git | MlFlow | scikit-learn | scipy | xgboost | LightGBM | streamlit

Data Scientist & Engineer PRIMAVERA BSS — Aug 2020 – May 2021 — Braga, Portugal

- Employed dual methodologies for component failure prediction: Regression to predict failure timing, and Classification to forecast failure within specific intervals.Created micro-services to capture and store data within Azure Data Lake. Developed data transformation and processing pipelines, primarily utilizing Azure Data Factory and Databricks. Established a software access point to integrate model outcomes.

Python | SQL | PySpark | Databricks | Azure | TFS | MlFlow | scikit-learn | scipy | xgboost | H2O | streamlit

Software Engineer PRIMAVERA BSS — Jun 2018 – May 2021 — Braga, Portugal

- Design and deploy a cloud-based maintenance software using .NET/.NET Core (C#), SQL, EntityFramework, and Angular. Oversee product life cycle management, ensuring Azure environment adherence to best practices, and implementing CI/CD pipelines.

Software Developer Pinto Brasil Group — Oct 2017 – Jun 2018 — Guimarães, Portugal

*Summer Internship Jun-Aug 2017 (concurrent with work at Idioma de Tons)

- Web solutions development (Idea Management Software, R&D product interfaces).
- R&D project: Prediction and identification of product failure moments and causes.

Textile Operator Idioma de Tons — May 2014 – Oct 2017 — Guimarães, Portugal

- Crucial for financial independence and university pursuit, shaping identity and fostering profound personal growth.

PUBLICATIONS

- Candeias, A., Silva, I., Sousa, V., & Marcelino, J. (2023). Tailor: Size recommendations for high-end fashion marketplaces. *16th ACM Conference on Recommender Systems - Fashion RecSys, 2023*. https://arxiv.org/abs/2401.01978.
- Ferreira, L., Pilastri, A., Sousa, V., Romano, F., & Cortez, P. (2021). Prediction of maintenance equipment failures using automated machine learning. *IDEAL 21*, 259–267. https:/link.springer.com/chapter/10.1007/978-3-030-91608-4_26.

EDUCATION

Engineering and Management of Information Systems (Integrated Master's) University of Minho

Sept 2015 – Dec 2020 **Thesis**: Artificial Intelligence for Predictive Maintenance (17/20)

LANGUAGES

Portuguese
English