Fran Valero

Applied Scientist · ML/NLP

Barcelona, Spain

☎ (+34) 678 812 474 | ✉ franbvalero@gmail.com | 🏠 franbvalero.github.io | franbvalero | franbvalero | @franbvalero | 🎓 Fran Valero

Work Experience

Amazon *Barcelona, Spain*

Applied Scientist *Apr. 2022 - Present*

- Applied ML/NLP to several community feedback (reviews and Q&A) projects to inform customer purchase decisions. The most impactful projects were:
    - AI-generated customer review highlights (see Amazon blog post)
    - Ranking text reviews
- Mentored a SDE team to increase the automation rate for automatically moderate different content types (text, image, and video), resulting in an annual cost reduction of MM$ in human annotations.

Amazon *Barcelona, Spain*

Data Scientist *Jun. 2021 - Mar, 2022*

- Worked on building a shopping queries dataset for improving the quality of search results. This work was presented at the Amazon Product Search competition as a part of KDD 2022 [1]

Deezer *Paris, France*

R&D Intern *Nov. 2020 - May. 2021*

- Worked on podcast topic modeling using textual metadata. This work was accepted as a full paper at ECIR 2022 [2]

Kustomer *Madrid, Spain*

ML/NLP Engineer Intern *June. 2020 - Sep. 2020*

- Worked on enhancing the intent capabilities in CRM chats fine-tuning pre-trained transformers models

Department of Software and Computing Systems, University of Alicante *Alicante, Spain*

Research Engineer *Apr. 2019 - Sep. 2019*

- Worked on quality estimation for machine translation systems using regression models

Department of Software and Computing Systems, University of Alicante *Alicante, Spain*

Research Intern *Jan. 2018, Mar. 2019*

- Worked on several projects in the area of text sequence labelling and machine translation using RNNs, Seq2Seq and cross-lingual word embeddings

Education

Polytechnic University of Valencia *Valencia, Spain*

Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging *Sep. 2019 - Sep. 2020*

- Master's Thesis: *Data augmentation for text classification* advised by Francisco Casacuberta

University of Alicante *Alicante, Spain*

Bachelor's Degree in Compute Science Engineering *Sep. 2014 - Feb. 2019*

- High Academic Performance Group
- Bachelor's Thesis: *Neural Part-of-Speech tagging* advised by Juan Antonio Pérez-Ortiz and Felipe Sánchez-Martínez

Publications

[1] Chandan K. Reddy et al. *Shopping Queries Dataset: A Large-Scale ESCI Benchmark for Improving Product Search*. 2022. arXiv: 2206.06588 [cs.IR]. URL: https://arxiv.org/abs/2206.06588.

[2] Francisco B. Valero, Marion Baranes, and Elena V. Epure. "Topic Modeling on Podcast Short-Text Metadata". In: *44th European Conference on Information Retrieval (ECIR)*. 2022.

Skills

Programming Languages	Python, LaTeX, Java, C/C++, bash, SQL
Technologies	PyTorch, PySpark, Transformers, scikit-learn, linux terminal, Git, Docker, AWS
Compute science	Data structures, algorithms and object-oriented software design
Natural Languages	Native in Spanish and professional working proficiency in English