Pablo de Castro Manzano
Curriculum Vitae

pablodecm@gmail.com
Madrid Area / Remote (based in Spain)

Profile

Scientist and Software Engineer passionate about technological innovation and entrepreneurship.
Committed to the practical application of knowledge to address critical challenges in society.
Broad technical knowledge (machine learning, data science and software/data/systems engineering).
Proven ability to understand cross-domain problems, proposing and building end-to-end solutions.
Demonstrated technical management and leadership skills in heterogenous and fast-paced environments.
Physics PhD with academic advanced/research-level expertise in Statistics and Machine Learning.
Expert in C++, Python and Web development, as well as data analysis abstractions and tools.
Business-savvy, with a strong acumen for product management and development.
Experienced in conveying complex technical concepts, teaching and mentoring others.
Excellent verbal and writing communication skills in English and Spanish.

Experience

2022-ongoing **Software Architect and Product Manager**, Reforestum (Startup - Climate Tech), Spain
Lead and coordinate the product management and development efforts through a pivot of the startup focus from a voluntary carbon market marketplace to an enterprise data platform.
Successfully plan and deploy frequent product iterations from start to finish, including integrations with customers and external data sources.
Contribute to the product and engineering efforts at many different levels (customer interviews/reseach, product design, software engineering, cloud infrastructure, etc)
Provide advice and support for the founding team through the strategic shift of the business direction.

2019-ongoing **Guest Lecturer**, Master of Data Science UC-UIMP-CSIC (Public University), Spain
Teaching each year a 3-day applied data analysis laboratory focussed on Data Science for the Internet of Things (IoT). More info and materials in   github.com/pablodecm/datalab_ml_iot.

2019-2022 **Head of Machine Learning Engineering**, Tree Technology (SME - IT & Data Consultancy), Spain
Technical leadership in for a cross-disciplinary chapter focussed on designing, building and improving solutions, products and processes with data science, machine learning and computer vision technologies both in the context of commercial clients and European R&D projects .
Responsibilities included: technical design and coordination of projects and proposals, codefinition of technological stack and methodologies, ensure and promote technical excellence and technology transfer, support and mentorship for other chapter members and collaboration in organization-wide initiatives.
Highlighted projects (majority in a technical lead role):
- *Industry Sector*: analyse, design and plan solutions for improving productivity and existing processes in several industry sectors.
- *Education Sector*: integrate and analyze automatic transcriptions using cloud services for different video-conference providers for search purposes.
- *DECODER H2020 R&D*: integrate natural language processing models for code summarization and variable misuse as services within a web application.
- *Internal Project*: iteratively improve the data, machine learning and software development methodologies and infrastructure in order to make technical teams more autonomous and productive.

Senior Data Scientist, Tree Technology (SME - IT & Data Consultancy), Spain
Work on a diverse set of projects that heavily rely on data science, machine Learning and computer vision technologies, both in the context of commercial clients that want innovative solutions and European R&D projects.
Highlighted projects (majority in a technical lead role):
- *Education Sector*: design, build and deploy into production a system for advanced video analytics using deep learning technologies and fully integrated with their cloud infrastructure.
- *Education Sector*: design, build and integrate heterogenous business data sources and provide a custom flexible solution to provide insights and visualizations to stakeholders.
- *Finance Sector*: evaluate the viability of a system for public companies default prediction using machine learning technologies.
- *Electric Power Sector*: design and build feasibility demonstrator of an end-to-end platform for the automatic detection of anomalies in electrical lines and towers using deep learning.
- *Robotics H2020 R&D*: design and advise in core data and machine and deep learning competencies in the projects, mainly in the context of the visual perception module.

2019-2021	**Technical Advisor and Contractor**, Reforestum (Startup - Climate Tech), Spain
	Co-design the data strategy and approach for a forest monitoring, verification and reporting system for reforestation and conservation projects, that supports one of the core value propositions of the startup (i.e. transparency).
	Create a prototype of a monitoring system for large forest conservation projects based on the use of machine and deep learning on satellite imagery.
	Co-design, execute and validate a plan for integrating a monitoring system for the project in the production web application of a carbon credit marketplace.
	Support and advice startup in various other technical domains such as software and data architecture and cloud infrastructure.
2015 - 2018	**MSCA Early Stage Researcher**, *INFN - Sezione di Padova* (Research Institute), Italy
	Work within the AMVA4NewPhysics H2020 project, whose aim is to develop and apply state of the art machine learning techniques for High Energy Physics data analyses.
	Main projects (all very focussed on the data analysis and machine learning side):
	- New machine learning technique to construct inference-aware summary statistics.
	- Non-resonant Higgs pair production analysis (bbbb channel) at the LHC with the CMS detector.
	- Integration of TensorFlow-based multi-class jet classifer model in CMS experiment software.
Winter 2016	**Academic Secondment**, *University California Irvine* (Research University), US
	Collaboration with researches at the UCI Center of Machine Learning on differentiable approximations of histograms to build inference-aware losses for neural networks and the role of new deep learning techniques on jet quark-gluon tagging using computer-vision techniques.
Autumn 2016	**Industrial Secondment**, *SDG Consulting Milan* (SME - IT & Data Consultancy), Italy
	Work on applications topological data analysis and developed a open-source package re-implementing the MAPPER algorithm with a scikit-learn-like API.
2014 - 2015	**Research Project Associate**, *University of Cantabria* (Public University), Spain
	Collaborate in data analyses within the CMS Collaboration.
Summer 2015	**Research Internship**, *Brown University* (Research University), US
	Carried out part of Master's thesis with the Experimental Particle Physics research group.
Summer 2014	**CERN Summer Student**, *CERN* (International Research Organization), Switzerland
	Work with an experimental research team on characterising silicon detectors using lasers. Developed an open-source simulator of the of drift dynamics of carrier distributions in complex semiconductor detectors.
Spring 2014	**Research Internship**, *Instituto de Física de Cantabria (IFCA)* (National Research Institute), Spain
	Focus on the use of ontologies, knowledge bases and semantic web technologies to design a system for data preservation in High Energy Physics.

Education

2015-2019	**PhD in Physics**, *University of Padua*, Italy
	Doctor Europaeus Cum Laude
	PhD Thesis: "Statistical Learning and Inference in Particle Collider Experiments"
	Available online at https://github.com/pablodecm/phd_thesis
2014-2015	**Master's Degree in Physics, Instrumentation and the Environment**, *University of Cantabria*, Spain
	Average grade: 9.7/10.0 - Specialty in Advanced Physics
	Master's thesis: " Measurement of CMS b-tagging efficiencies using the Flavour-tag Consistency Method at a center-of-mass-energy of 13 TeV "
2010-2014	**Bachelor's Degree in Physics (4 years)**, *University of Cantabria*, Spain
	Average grade: 8.6/10.0 - Mention in Fundamental Physics
	Final Year Project: " Measurement of the W+W- production cross section in pp collisions at a center-of-mass energy of 8 TeV "
2012-2013	**Physics Exchange Student**, *Imperial College London*, UK
	1st Class (70% GPA)
2008-2010	**Spanish Baccalaureate in Science and International Baccalaureate**, *I.E.S. Santa Clara*, Spain
	University Access Qualification: 12.1/14.0

Skills

Languages	**Spanish**: Native speaker
	English: Proficient user (> **C1 level**) with the following certifications:
	- Cambridge Advance English (**CAE**): Grade B (June 2013)
	- Test of English as a Foreign Language (**TOEFL**): 101/120 Score (December 2013)
	Very experienced technical writer in English and Spanish.
	Basic knowledge of Italian and Serbian (no certifications).

Computing

Advanced Linux and Unix system administrator (>10 years)

DevOps Expertise: version control, testing, security, code review, CI/CD, containerization, IaC and other methodologies.

Programming Languages: many projects carried out using Python, C++ and Javascript among others with modern style and practices.

Software and Data Architectures: familiar with common design patterns and methodologies and useful data structures both at high and low level.

Web Development: knowledge of common backend and frontend technologies (e.g. Node, Flask, FastAPI, React, etc), micro-services architectures and API-First design or standards (e.g. OpenAPI) and several database technologies (both SQL and NoSQL).

Containers and Cloud: expertise with container technologies for development and production (Docker, Docker Compose, Kubernetes, Helm) and core services of main cloud providers (AWS, Azure, Google Cloud and Digital Ocean) for compute and storage.

Data Engineering: understanding of core data engineering and big data abstractions such a data lake, data warehouse, ETLs, map-reduce, batch processing and real-time data pipelines.

Data Analysis: expertise with common analysis libraries and frameworks such as numpy, pandas, TensorFlow, PyTorch, scikit-learn, ROOT, R and many more.

Data Visualization: used to common libraries such as matplotlib, ggplot, Plotly and D3.js.

Scientific/technical document creation with Latex/Markdown

Data Science and Machine Learning

Extensive experience with diverse data manipulation, management, visualization and exploratory data analysis techniques and tools: NumPy, Pandas, Matplotlib, JupyterLab, Plotly, PySpark, SQL, Elasticsearch, Kibana among others.

End-to-end design and implementation of supervised machine learning systems for heterogeneous data types (tabular, text, image, video, point clouds or graphs) either using custom implementations or different libraries such as PyTorch, xgboost, TensorFlow, scikit-learn, etc.

  - Knowledge of a large breadth of supervised techniques and its different range of applications as a function of the data types (neural networks, nearest neighbors, gaussian processes, naive bayes, random forest, gradient boosting algorithms, ensembling).

  - Deep expertise with deep neural network technologies and architectures (including self-supervision, encoder-decoder, convolutional, recurrent, transformers, adversarial setups, etc.) applied to different tasks (classification, regression, image segmentation and object detection, image or text generation) and types of data.

Hands-on experience with the design, architecture, orchestration and deployment of data and machine learning training and production pipelines on premises and on public clouds with different tools such as Kubeflow, AWS Batch, Prefect, Airflow, Dask among others.

Familiarity with unsupervised techniques for dimensionality reduction and clustering, recommender systems and reinforcement learning abstractions.

Leadership and Management

Efficient communication with both business and technical stakeholders.

Technical team coordination and mentoring.

Project/product/process ownership all the way from problem identification to solution ideation and validation.

Familiar with Lean and Agile methodologies and common tools for project planning and management.

Awards and Grants

2015-2018	**Marie Sklodowska-Curie ESR fellowship**, *AMVA4NewPhysics ITN*, EU
2015	**Brown University Exchange Scholarship**, *University of Cantabria*, Spain
2014	**CERN Summer Student**, *CERN*, Switzerland
2013-2014	**Undergraduate Research Scholarship**, *Spanish Government*, Spain
2012-2013	**Erasmus Scholarship with Excellence Mention**, *Spanish Government*, Spain

Selected Publications

Author of 200+ publications as a member of the CMS Collaboration. See   Google Scholar Profile for full list.

Here is a list of selected non-collaboration publications and a subset of collaboration publications with major individual contributions:

Book Chapter	"Dealing with Nuisance Parameters". T. Dorigo and P. de Castro. Artificial Intelligence for High Energy Physics. World Scientific (2021). doi:10.1142/12200.
Journal Paper	"INFERNO: Inference-Aware Neural Optimisation". P. de Castro and T. Dorigo. Computer Physics Communications (2019). doi:10.1016/j.cpc.2019.06.007.
Journal Paper	"Search for nonresonant Higgs boson pair production in the bbbb final state at 13 TeV". CMS Collaboration. JHEP 04 (2019) 112. doi:10.1007/JHEP04(2019)112.
Journal Paper	"Combination of searches for Higgs boson pair production in proton-proton collisions at 13 TeV". CMS Collaboration. Phys. Rev. Lett. 122, 121803 (2019). doi:10.1103/PhysRevLett.122.121803.
Journal Paper	"TRACS: A multi-thread transient current simulator for micro strips and pad detectors". J. Calvo, P. de Castro et al. Nucl. Instrum. Methods Phys (2019). doi:10.1016/j.nima.2018.11.132.
Workshop Paper	"DeepJet: Generic physics object based jet multi-class classification for LHC experiments". NeurIPS 2017 DSPS Workshop. Markus Stoye et al on behalf of CMS Collaboration. December 2017. Workshop Paper.

Selected Talks and Posters

Invited Talk	"INFERNO: Inference-Aware Neural Optimisation" **Likelihood-Free Inference Workshop**, March 2019, Flatiron Institute, New York
Conference Talk	"Reducing the impact of systematic uncertainties with inference-aware summary statistics" **Advanced Computing and Analysis Techniques in Physics Research**, March 2019, Sans-Fee, Switzerland
Poster	"INFERNO: Inference-Aware Neural Optimisation" **Advanced Statistics for Physics Discovery**, September 2018, Padua, Italy
Talk and Poster	"Direct Learning of Systematics-Aware Summary Statistics" (*awarded best poster prize*) **XIIIth Quark Confinement and the Hadron Spectrum**, August 2018, Maynooth, Ireland
Workshop Talk	"Direct Learning of Systematics-Aware Summary Statistics". **2nd Inter-experimental Machine Learning Working Group Workshop**, April 2018, CERN, Switzerland
Workshop Talk	"TRACS: Transient Current Simulator" **25th RD50 Workshop on Radiation hard semiconductor devices for very high luminosity colliders**, CERN, Switzerland

Other Experience and Certifications

Hackathon Winner	**Technical Challenge: Cal.com Best Message Queue Implementation** December 2023, *OSSHack NYC at Cornell Tech*, US
Teaching	**Datalab - Practical Data Science for IoT** 2019-2023 Master of Data Science (jointly organized by UC-UIMP-CSIC), Santander, Spain
Public Seminar	**Adapting Machine Learning for Scientific Discovery** (in Spanish) March 2017, University of Oviedo, Spain
Science Outreach	**Written recurrently in AMVA4NewPhysics outreach blog** 2015-2018, AMVA4NewPhysics H2020 Project Blog

⚫ pablodecm ⚫ pablodecm ⚫ pablodecm ⚫ https://pablodecm.com/

January 30, 2024