Mattias Arro

mattias.arro@gmail.com

- **Machine learning engineer with 10+ years of software engineering experience.**
- Strong theoretical understanding of machine learning, statistics, and large-scale data mining.
- 2.5+ years of commercial experience with building end-to-end ML solutions, including
    - deep learning models
    - scalable big data processing pipelines
    - recommender engines
    - interactive visualizations.
- Obsessed with code quality, developing test-driven when possible.
- Up-to-date on the state of the art in ML and DL research.

KTH Royal Institute of Technology (2016 - 2017)   Sweden

MSc Data Science (year 2 of a 2–year program)   *Cum laude*
Thesis: **Label-efficient multi-objective machine learning for e-commerce.**   *Grade: A*
    Exploration of transfer, multi-objective and active learning across shallow   *(100%)*
    and deep neural network architectures for e-commerce product classification.
Topics: large-scale machine learning, deep learning, data mining, recommender systems.

Eindhoven University of Technology (2015 - 2016)   The Netherlands

MSc Data Science (year 1 of a 2–year program)   *Cum laude*
Topics: statistics, statistical learning theory, advanced algorithms (proofs & analysis),   *Grades: 8.16*
    IoT, data mining, process mining, information retrieval, entrepreneurship.   *(8 = 85%)*

University of Southampton (2010 - 2013)   United Kingdom

BSc Computer Science   *2:1 honours*
Thesis: **Exploratory Web Analytics.**   *Grade: A+*
    Engineering Approaches from Server-Side TDD to Client-Side Rendering   *(80%)*
Topics: compilers, theory of computing, mathematics (discrete maths, linear algebra, calculus,
    probability), software engineering, programming (OOP, scripting languages), formal methods.

Programming languages:	**Python**, **Clojure**, **JavaScript**, R, Scala, Ruby, Java, C
Machine learning:	**Dask**, scikit-learn, XGboost, pandas, numpy, scipy, Prediction.IO
Deep learning:	**TensorFlow**, Keras, mxnet, tensor2tensor; GPU training
Big Data:	**Apache Beam, Apache Spark**, Hadoop, TensorFlow Transform
Queues & Orchestration:	**RabbitMQ**, GCP Pubsub, Logstash; **Airflow**, **subdag**, MLflow; Celery
Databases:	**ElasticSearch**, MongoDB, DynamoDB, MySQL/PostgreSQL/CloudSQL
Front-end & visualisation:	**D3.js**, **matplotlib**, Reagent, React, Angular, Ember, Bootstrap, jQuery
Back-end web development:	**Flask**, Meteor, Ruby on Rails, Node.js, Sinatra
Testing:	**py.test**, Rspec, Cucumber, Selenium
Cloud Platforms:	GCP: **ML Engine**, **Dataflow**, Dataprep, GCS; AWS: EC2, S3; Heroku
OSs & VMs:	macOS, **Linux** (Debian); **Kubernetes**, Docker, Vagrant
VCS, dependencies, PM:	**git**, Tower; pip, brew, pyenv, conda, lein, gem; Trello, Pivotal
Text Editors & IDEs:	Atom, RStudio, vim, PyCharm, IntelliJ IDEA, LaTeX

Data science contracts & internships

Shareight Ltd (octer.co.uk)
Data scientist / ML engineer (contract via Subspace AI Ltd)

London
Nov 2017 – Nov 2018

Building an end-to-end ML pipeline for product categorisation and visual similarity. Employing multi-objective, multi-modality, active and transfer learning approaches. Training the team to use the built ML model independently. Using GCP stack (Dataflow, ML Engine, PubSub, Dataprep, GCS, CloudSQL) and OSS (TensorFlow, Airflow, Apache Beam, ElasticSearch, RabbitMQ, Flask, React).

Lingvist Technologies OÜ (lingvist.com)
Data scientist / ML engineer (contract via Apparat Cloud OÜ)

Tallinn / London
Jul 2017 - Nov 2017

Building generative and discriminative neural NLP model for semisupervised learning and content generation. Using TensorFlow, tensor2tensor, Keras.

Rise Art Ltd (riseart.com)
Data scientist / ML engineer (contract via Apparat Cloud OÜ)

London / remote
Jun 2016 - Jun 2017

Building a recommender engine based on deep learning and matrix factorization using Python, TensorFlow, Spark and Google Cloud ML Engine. Implementing and measuring the performance of various recommender algorithms. End-to-end engineering.

RISE SICS (ri.se)
Research intern

Stockholm
Nov 2016 – Apr 2017

Working on my MSc thesis; attending workshops at the Decisions, Networks and Analytics lab.

Teleport Inc (teleport.org)
Data scientist (summer intern)

Tallinn
Jun – Jul 2015

Creating the New York Teleport based on the Bay Area code: refactoring Python code, implementing life quality metrics calculations (transitioning from R to Python).

Short term work:
- TeamScope.io
- YouModels.com

Awards:
- Winner of BDL Accelerate 2016 (Garage48)
- Kristjan Jaak scholarship (Archimedes)

Theses:
- MSc: http://bit.ly/2NaOXQY
- BSc: http://bit.ly/2RaFgVW

Data science theory

deep neural networks (DNN, ANN), convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, variational autoencoders (VAE), random forests, gradient boosted trees, support vector machines, matrix factorisation / dimensionality reduction (SVD, ALS), PCA, bayes networks, probabilistic graphical models, regularisation techniques, linear regression, logistic regression

Influential MOOCs:
- *Introduction to Probability* (MIT)
- *The Analytics Edge* (MIT)
- *Machine Learning* (Stanford)
- *Neural Networks for Machine Learning* (Toronto)
- Mining of Massive Datasets (Stanford)

Web development contracts & other involvement

Subspace AI Ltd
London
Founder, full-stack ML engineer, data scientist
Feb 2018 - ...
Running a data science consultancy: offering full stack ML engineering, model building and training.

Apparat Cloud OÜ
UK / Tallinn
Founder, full-stack developer
2008 - 2017
Running a web development consultancy: offering full-stack web development for small companies and start-ups using Rails and different front end frameworks.

asmallworld.net
UK / Tallinn
Full-stack web developer using Rails (contract via Apparat Cloud OÜ)
May 2014 – May 2015
Implementing features for the social network with a substantial code base.

likette.com
UK
Rails full-stack developer (contract via Apparat Cloud OÜ)
Oct 2010–Feb 2012
Architecture & full-stack development. Subcontracting design and some front-end development.

Credit Suisse
London
Technical analyst (summer intern)
2012

Freelance work
Tallinn
PHP full-stack developer
2005 - 2008
Started off as a hobbyist PHP developer in 2003, built around a dozen custom CMS solutions, WordPress installations and minor e-commerce applications for small companies.

Open source projects

subdag
2018
A minimal task manager for ML workflows in Python.
https://github.com/subspace-ai/subdag

Conditional Variational Autoencoder (seq2seq model)
2017
TensorFlow implementation of "Toward Controlled Generation of Text" (Hu et al. 2017)
https://github.com/mattiasarro/seq2seq-cvae-tensorflow

World Bank dataset visualisation
2015
D3.js and Angular implementation of World Bank data on Internet usage across countries.
https://github.com/mattiasarro/world-bank-visualization

statt web analytics
2013
SaaS web analytics platform that preserves and visualises event-level data using D3.js, Ember.js, Rails.
https://github.com/mattiasarro/statt-display
https://github.com/mattiasarro/statt-collect

draggable date selector for jQuery calendar
2013
https://github.com/mattiasarro/jquery-ui-daterange-draggable

apparat CMS
2012
A generic website CMS built with Rails (discontinued). Used for running several websites (2008 - 2012).