

Mattias Arro

mattias.arro@gmail.com

- **Machine learning engineer with 10+ years of software engineering experience.**
- 4+ years of commercial experience with building end-to-end ML solutions, including
 - deep learning, NLP, recommender systems
 - scalable big data pipelines
 - continuous model train/deployment and experiment management systems.
- Focused on code quality & automation, careful not to over-engineer.
- 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)

Thesis: **Label-efficient multi-objective machine learning for e-commerce.**

Exploration of transfer, multi-objective and active learning across shallow and deep neural network architectures for e-commerce product classification.

Topics: large-scale machine learning, deep learning, data mining, recommender systems.

Grade: A
(100%)

Eindhoven University of Technology (2015 - 2016)

The Netherlands

MSc Data Science (year 1 of a 2-year program)

Topics: statistics, statistical learning theory, advanced algorithms (proofs & analysis), data mining, process mining, information retrieval, entrepreneurship.

Cum laude
Grades: 8.16
(8 = 85%)

University of Southampton (2010 - 2013)

United Kingdom

BSc Computer Science

Thesis: **Exploratory Web Analytics**

Engineering Approaches from Server-Side TDD to Client-Side Rendering

Topics: compilers, theory of computing, mathematics (discrete maths, linear algebra, calculus, probability), software engineering, programming, formal methods.

2:1 honours
Grade: A+
(80%)

Programming languages:	Python, Clojure, JavaScript , R, Scala, Ruby, Java, C
Machine learning:	Dask, pandas, numpy , scipy, scikit-learn, XGboost
Deep learning:	TensorFlow, PyTorch , spacy, Keras, mxnet; GPU training
Big Data:	Apache Beam, Apache Spark , petastorm, TF Transform, onyx (Clojure)
Databases:	ElasticSearch, SQL , MongoDB, DynamoDB
Queues:	RabbitMQ , GCP Pubsub, Logstash
Workflows & MLOps:	Kubeflow, Kubeflow Pipelines, Airflow , Celery
DevOps:	Kubernetes, Docker, Linux , Terraform, CircleCI
Cloud Platforms:	<u>AWS</u> : EKS, EC2 , VPC; <u>GCP</u> : ML Engine, Dataflow , Dataprep
Testing:	py.test , Rspec, Cucumber, Selenium
Front-end & visualisation:	D3.js, matplotlib , Reagent, React, Angular, Ember, Bootstrap, jQuery
Back-end web development:	Flask , Meteor, Ruby on Rails, Node.js, Sinatra
VCS, dependencies, PM:	git , git-flow; pip, brew, pyenv, conda, lein, gem; Trello, Pivotal, Jira

Data science contracts

FunnelAI - funnelai.com

ML / NLP engineer (contract via Subspace AI Ltd)

London
July 2019 – ...

Built a continuous model retraining, tuning & deployment system on Kubeflow. The design emphasises DRY, reusable components, scalable data processing, and ease of adding additional models and training objectives. Performed multi-objective training on top of SotA NLP models (BERT, ROBERTA).

Factmata Ltd - factmata.com

ML / NLP engineer (contract via Subspace AI Ltd)

London
Jan 2019 – July 2019

Designed and lead the development of a multi-stage NLP pipeline (on Kubeflow) for a media monitoring tool, involving clustering, semantic similarity, keyphrase extraction, etc. Designed and partially implemented a continuous model retraining, tuning & deployment system with active learning.

Shareight Ltd - octer.co.uk

ML engineer / data scientist (contract via Subspace AI Ltd)

London
Nov 2017 – Nov 2018

Built an end-to-end ML pipeline for product categorisation and visual similarity. Employed multi-objective, multi-modality, active and transfer learning approaches. Coached the team to use the ML system independently.

Lingvist Technologies OÜ - lingvist.com

Deep learning engineer (contract via Apparat Cloud OÜ)

Tallinn / London
July 2017 - Nov 2017

Built a generative / discriminative neural NLP model for semisupervised learning and content generation.

Rise Art Ltd - riseart.com

ML engineer / data scientist (contract via Apparat Cloud OÜ)

London / remote
June 2016 - June 2017

Built a recommender engine based on deep learning and matrix factorization. Implemented and measured the performance of various recommender algorithms. End-to-end engineering.

Teleport Inc - teleport.org

Data scientist (summer intern)

Tallinn
June – July 2015

Contributed to the New York Teleport app: refactoring and transitioning code from R to Python.

Talks given:

- Data Natives (Berlin 11.2019)
- EGG (London 07.2019)
- AI WTF (London 06.2019)

Awards:

- Winner of BDL Accelerate (2016)
- Kristjan Jaak scholarship (2015)

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:

- *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

Founder, full-stack ML engineer, data scientist

Running a data science consultancy: offering full stack ML engineering, model building and training.

London

Feb 2018 - ...

Apparat Cloud OÜ

Founder, full-stack developer

Ran a web development consultancy which offered full-stack web development for small companies and start-ups, using Rails and different front end frameworks.

UK / Tallinn

2008 - 2017

asmallworld.net

Full-stack web developer using Rails (contract via Apparat Cloud OÜ)

Implemented features for the social network with a substantial code base.

UK / Tallinn

May 2014 – May 2015

likette.com

Rails full-stack developer (contract via Apparat Cloud OÜ)

Architecture & full-stack development.

UK

Oct 2010 – Feb 2012

Credit Suisse

Technical analyst (summer intern)

Back office, Java development.

London

2012

Freelance work

PHP full-stack developer

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.

Tallinn

2005 - 2008

Open source projects

subdag

A minimal task manager for ML workflows in Python.

<https://github.com/subspace-ai/subdag>

2018

Conditional Variational Autoencoder (seq2seq model)

TensorFlow implementation of "Toward Controlled Generation of Text" (Hu et al. 2017)

<https://github.com/mattiasarro/seq2seq-cvae-tensorflow>

2017

World Bank dataset visualisation

D3.js and Angular implementation of World Bank data on Internet usage across countries.

<https://github.com/mattiasarro/world-bank-visualization>

2015

statt web analytics

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>

2013

draggable date selector for jQuery calendar

<https://github.com/mattiasarro/jquery-ui-daterange-draggable>

2013

apparat CMS

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

2012