

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.
- 3+ 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 and deep neural network architectures for e-commerce product classification.

(100%)

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), IoT, data mining, process mining, information retrieval, entrepreneurship.

Grades: 8.16

(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, ClojureScript, 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, 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 PROJECTS

Shareight Ltd (octer.co.uk)

Data scientist / ML engineer

London
Nov 2017 – ...

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

Estonia / UK
Jul 2017 - Nov 2017

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

Rise Art Ltd (riseart.com)

Data scientist / ML engineer

UK / remote
Jun 2016 - Jun 2017

Building a recommender engine as an ensemble of deep learning models 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

Sweden
Nov 2016 – Apr 2017

Attending workshops at the Decisions, Networks and Analytics lab while working on personal projects.

Teleport Inc (teleport.org)

Data scientist (summer intern)

Estonia
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 University)
- *Neural Networks for Machine Learning* (Toronto)

OTHER PROJECTS

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

Full-stack web development for small companies and start-ups using Rails and different front end frameworks.

UK / Estonia

2008 - 2017

asmallworld.net

Full-stack web developer (Rails)

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

remote

May 2014 – May 2015

likette.com

Rails full-stack developer

Architecture & full-stack development. Subcontracting design and some front-end development.

UK

Oct 2010–Feb 2012

Credit Suisse

Technical analyst (summer intern)

UK

2012

e-Laen.ee (no longer operates)

Rails full-stack developer

Contract work for an online microfinance company. Refactoring to REST, maintenance.

Estonia

2009

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.

Estonia

2005 - 2008

OPEN SOURCE PROJECTS

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