

ROBERT ALONSO ADUVIRI CHOQUE

Pontificia Universidad Católica del Perú
Lima, Peru
robert.aduviri@pucp.edu.pe

robert-aduviri.github.io
linkedin.com/in/robert-aduviri
github.com/robert-aduviri
@robert_aduviri

EDUCATION

Massachusetts Institute of Technology (MITx on edX) **May 2019 - November 2020**
MITx MicroMasters Program in Statistics and Data Science *Capstone Exam Grade: A*

Comprehends subjects of the first year of the MIT PhD Program in Social & Engineering Systems: Probabilistic Models, Statistical Inference, Data Analysis & Experimental Design in Social Science, and Machine Learning & Deep Learning
Full scholarship funded by Aporta Institute for Advanced Analytics and Data Science.

Pontificia Universidad Católica del Perú (PUCP) **March 2013 - July 2018**
B.Sc. in Informatics Engineering, Faculty of Science and Engineering *Cumulative GPA: 4.0/4.0*

Academic Distinction: Valedictorian (rank 1 among 289 graduates across all the faculty)
Full tuition funded by the University Academic Excellence Scholarship (BEA)
Thesis research project: *Multiobjective genetic algorithm for the optimization of humanitarian relief supply distribution in Peru*. Obtained the PADET award, granted to the best undergraduate thesis research projects (\$5,000 PEN)

University of Calgary (U of C, Canada) **January 2016 - April 2016**
Computer Science, Faculty of Science *Term GPA: 4.0/4.0*

International academic exchange term funded by the Emerging Leaders in the Americas Program (ELAP) Scholarship, awarded by the Government of Canada: \$7,200 CAD financial support

Universidade de São Paulo (USP, Brazil) **August 2015 - December 2015**
Ciência da Computação, Instituto de Matemática e Estatística *Term Average: 8.2/10*

International academic exchange term funded by the International Student Mobility Scholarship (PUCP): \$4,000 USD financial support

WORK EXPERIENCE

Kanto **December 2017 - Present**
Data Scientist - Software Engineer

- Worked closely with the founding team and marketing & UX team in order to define unit economics and customer journey KPIs, automating its reporting through dashboards in Google Data Studio. Implementing and deploying ETL pipelines in GCP (Cloud Storage, BigQuery, Dataflow) and AWS (S3, Redshift, Athena, Glue), integrating information from multiple sources of data such as PostgreSQL, Mixpanel Analytics and Elasticsearch databases.
- Created forecast models for establishing KPIs baselines, aligned with business strategies impact levers, in order to estimate and optimize business KPIs objectives through systematic experimentation (e.g. A/B testing) and user research.
- Worked on the back-end codebase and infrastructure, based on Django REST and deployed on AWS Elastic Beanstalk and Lambda, implementing and maintaining web services and core algorithms for search, recommendations, user recordings & audio synchronization, singing voice pitch scoring (FFT & DTW), and optical character recognition.

- Proposed, researched, implemented and deployed a song recommender system based on matrix factorization with implicit-feedback [paper] from over three years of user interactions data, used for search and song recommendation chaining.
- Refactored and improved the karaoke videos lyrics extraction existing pipeline with Tesseract-OCR and the EAST convolutional neural network [paper] implemented in TensorFlow, containerizing it with Docker for distributed cloud processing on Microsoft Azure.

Machine Learning Tokyo

June 2020 - Present

Backend Software Engineer - Open Source Contributor

- Designed and implemented the infrastructure of the serverless API of the ML Search project, using AWS API Gateway, Lambda, DynamoDB and Elasticsearch.
- Set up the API development workflow on Flask and Zappa, together with the CI/CD DevOps processes with GitHub Actions.
- Maintainer of the API repository, providing code reviews and managing issues & pull requests.

Artificial Intelligence Research Group, PUCP

May 2017 - December 2019

Research Assistant - Software Engineer

- ***Towards disentangled representations via variational sparse coding:*** Reproduced the “*Variational Sparse Coding*” paper [link] in PyTorch and published a reproducibility report [paper] for the ICLR 2019 Reproducibility Challenge [peer reviews][reviewers] in the ReScience C journal. Presented improvements and extensions of the original work as an oral presentation and poster session at the ICML 2019 LatinX in AI workshop [oral presentation][code][slides][poster][interview].
- ***Feature selection algorithm recommendation for gene expression data with meta learning:*** Proposed and implemented matrix factorization and learning-to-rank approaches for metalearning of feature selection and classification tasks over genomic datasets. Published in the proceedings of the IEEE International Conference on Bioinformatics and Biomedicine 2018 [paper]. Presented at the poster session of the NeurIPS 2019 LatinX in AI workshop.
- ***Deep Learning for harmonized commodity description and coding system automatic classification:*** Implemented & deployed a bi-LSTM neural network model in PyTorch and a fastText model [paper] as a REST API on Flask and Gunicorn for large-scale product text classification, for an industry-collaboration project with FedEx Crossborder.
- ***Low-resource neural machine translation for a Peruvian minority language:*** Implemented a sequence-to-sequence LSTM neural network model in PyTorch for low-resource neural machine translation experiments [paper] and conducted transfer learning experiments

Pontificia Universidad Católica del Perú

January 2017 - December 2019

Teaching Assistant

- Courses: Machine Learning, Data Analysis, Artificial Intelligence, Applications of Computer Science.
- Assisted and evaluated up to 40 students during each laboratory session and presented specific topics during some lectures. Prepared boilerplate code for laboratory sessions, exams and quiz solutions.
- Conducted an 8-week practical course on Data Science with Python, covering the end-to-end Data Science process together with applications of Machine Learning & Deep Learning and Kaggle InClass competitions.

PUBLICATIONS

[Re] Variational Sparse Coding (May, 2019)

Alfredo De la Fuente and Robert Aduviri

ReScience C Journal, Volume 5 (2019)

DOI: 10.5281/zenodo.3161734

Feature selection algorithm recommendation for gene expression data through gradient boosting and neural network metamodels (December 2018)

Robert Aduviri, Daniel Matos and Edwin Villanueva

2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

DOI: 10.1109/BIBM.2018.8621397

COMPETITIONS, ACADEMIC CONFERENCES AND SUMMER SCHOOLS

July 2020 - LxMLS 2020 (formerly Lisbon, Portugal): Joined the Lisbon Machine Learning Summer School monitors team assisting the participants during the lectures and labs.

July 2020 - ACL 2020 (formerly Seattle, Washington): The largest conference on Computational Linguistics. Participated as a volunteer in the development of the virtual conference web platform on GitHub.

December 2019 - LatinX in AI Workshop at NeurIPS 2019 (Vancouver, Canada): The largest conference on Machine Learning & Neural Networks. Presented “*Feature selection algorithm recommendation for gene expression data with meta learning*” in the poster session.

July 2019 - LxMLS 2019 (Lisbon, Portugal): The Lisbon Machine Learning Summer School. Presented “*Towards disentangled representations via variational sparse coding*” in the poster session.

June 2019 - LatinX in AI Workshop at ICML 2019 (Long Beach, California): The largest conference on Machine Learning. Presented “*Towards disentangled representations via variational sparse coding*” in the poster session and as a highlight oral presentation.

December 2018 - IEEE International Conference on Bioinformatics and Biomedicine (Madrid, Spain): Presented “*Feature selection algorithm recommendation for gene expression data through gradient boosting and neural network metamodels*” in the poster session.

October 2018 - Winner, Movistar DataTalent - Kaggle (Lima, Peru): Winner of the data science competition hosted on Kaggle. \$3,000 USD prize.

September 2018 - Finalist, Data Science Game 2018 Finals - Kaggle (Paris, France): Qualified as part of the top 20 among 150 teams and 600 students from 30 countries to the finals in Paris [[leaderboard](#)] [[competition](#)].

July 2017 - The Yamagata-Andes Nations’ Double Triangle Program (Yamagata & Tokyo, Japan): Three-week short-term exchange program funded by the JASSO Scholarship awarded by the Government of Japan (\$80,000 JPY + travel & accommodation expenses)

June 2017 - York CVR-VISTA Vision Science Summer School (Toronto, Canada): Scholarship granted by the Centre for Vision Research (CVR) as part of the Vision: Science to Applications (VISTA) initiative organized by York University.

November 2016 - Runner-up, ACM-ICPC South America/South Finals (Lima, Peru): 2nd place in Peru and 4th place in the South America/South region among 144 teams and 432 students from 6 countries. Last phase before World Finals.

July 2016 - São Paulo Summer School of Advanced Science on Algorithms, Combinatorics and Optimization (São Paulo, Brazil): Presented “*An evolutionary optimization approach for disaster relief supplies distribution*” in the poster session.

March 2016 - Runner-up, Calgary Collegiate Programming Contest - Division 1 (Calgary, Canada): 2nd place in Calgary - first division. \$750 CAD prize.

SKILLS

Programming Languages
Software Development

Python, JavaScript, C++, Java, SQL, AMPL
Django REST, Flask, AWS Elastic Beanstalk, Lambda, DynamoDB,
S3, Redshift, Athena, Glue, GCP Cloud Storage, BigQuery, Dataflow,
Data Studio, FastAPI, Vue JS, React JS, Firebase, OpenGL

ML & Data Science
Languages

PyTorch, TensorFlow, Keras, Fastai, Spacy, Dash, Gensim, Scrapy
Native Spanish, Fluent English (TOEFL iBT [2015]: 104/120),
Intermediate Portuguese, Basic Japanese (JLPT N5 [2017]: 94.3/100)