Dorsa M. Arezooji

Data Scientist

NW8 9EL London, UK | d.arezooji@gmail.com | https://dorsa-arezooji.github.io | https://linkedin.com/in/dorsa-arezooji

EDUCATION

M.Sc. in Big Data Science | **Queen Mary** University of London, UK 09/2019 – 09/2020

- Graduating with **distinction** [**90/100**]
- Earned **distinction** in all 8 taken **modules** and the **final project**
- Recipient of the Queen Mary University of London **International Science and Engineering Excellence Award** in 2019
- Selected **Modules**:
  Big Data Processing [**93.8**] | Cloud Computing [**88.9**] | Machine Learning [**91.4**] | Data Mining [**97.2**] |
  Bayesian Decision & Risk Analysis [**95.7**] | Data Analytics [**88.9**] | Applied Statistics [**89.2**] | M.Sc Project [**85.7**] |

B.Sc. in Mechanical Engineering (Mechatronics) | Shahid Beheshti University, Tehran, Iran 09/2014 – 09/2018

- Graduated with **First Class Honors [1:1]**
- Selected Modules:
  Fundamentals of Programming (**C++**) | Project Management | Signal Processing | Modern Control Theory & Simulation |

PROJECTS

- **PortoDash**
  - Created and deployed a **cloud-based** investment **portfolio dashboard application** using **R**, **Shiny** and **HTML** in 1 week
  - Implemented several **risk** and **return models** along with an **interactive UI** to **visualize** assets and trends to gain insights
  - Added a feature to allow the user to easily **build** and **optimize** their **portfolio** by visualizing it & the **efficient frontier**
  - Used **machine learning,** and **Bayesian learning** (MCMC) to provide highly accurate (>96%) **market predictions**
- **Forex Rate Forecast with MCMC**
  - Used **Bayesian** structural time-series (bsts) in **R** to **predict forex** rates in **long-term** with a high accuracy of **98%**
  - **Visualized** the performance of the bsts model compared with widely used time-series methods like ARIMA (75% accuracy)
- **Forex API**
  - Created a **dynamic REST API** using **python** & **Flask**, designed a web **UI** and deployed it on **AWS** using **Docker** in 15 days
  - Implemented access to **Forex** data and created a **trading journal** service using a **Cassandra** database to calculate profits
  - Implemented market prediction, **API** keys, password **hashing**, and load balancing (**Kubernetes**) to improve functionality
- **FinDL**
  - Implemented **deep learning** methods (**LSTM**, **DNN**, **RNN**) in **python** to predict financial time-series with **96%** accuracy
  - Created a **Bayesian dynamic linear model** and compared its performance to those of the mentioned DL models
- **Retail Analytics**
  - Implemented **feature engineering** & **machine learning** algorithms including random forests and time-series analysis in **python** to **predict** sales with an accuracy of **97%**
  - Used Bayesian structure learning to discover causal links between **macro-economic** factors and sales with 72% precision
- **Ethereum Analysis**
  - Performed **big data** analysis of Ethereum contracts & transactions using **PySpark** & **Hadoop** to discover trends & insight
  - Analyzed scammers' graph, using **GraphFrames** to find 63 wallets used to accumulate stolen Ether
  - **Visualized** the trends and insights to better convey the results of the analysis
- **Natural Language Processing (NLP)**
  - Implemented naive Bayes classifiers, vector spaces, and locality sensitive hashing in **python** to extract **sentiment** from tweets using **nltk**
- **Dose-Response Modeling with MCMC**
  - Modelled the effects of a drug in a **probabilistic** framework to find the **optimal** dosage by creating a **hierarchical logistic regression model** using **R**, **stan** and **Hamiltonian Monte-Carlo** sampling with **95%** accuracy
  - **Visualized** the experimental data and the proposed model's dose-response curve using 3 types of prior distributions
  - Compared the **performance** of **MCMC algorithms** in R stan, pymc, and AgenaRisk using the same hierarchical model

EXPERIENCE

- **Freelance Data Scientist** | Freelancer.com, London, UK 06/2019 – Present
  - Provided a range of services including: **R** & **Python** Development, Data Engineering, Statistical Analysis, Machine Learning
  - Completed **100%** of project on time and within budget with **4.6/5** review scores | Profile Link
- **Technology Intern - IEUK** | Bright Network, London, UK 06/2020 – 07/2020
  - **Researched** the current market for facial authentication and machine learning algorithms
  - Wrote **user stories**, created a **project management** plan and a **communication** plan in 3 days
  - Attended technical seminars by Google, Amazon, Goldman Sachs, Accenture & Vodafone
- **R&D Engineer** | PATRA Vision, Tehran, Iran 03/2019 – 09/2019
  - **Optimized** the mechanical structure of Anea3D using FEM analysis, reducing the wight by **~ 35%** & saving up on costs
  - **Led a small team** of 3 assigned to the Anea3D redesign project in a start-up environment
- **Research Assistant** | Institute for Cognitive & Brain Sciences, Shahid Beheshti University, Iran 06/2018 – 07/2019
  - Wrote code in **Matlab** to **automate** signal processing, fractal analysis and machine learning (**SVM**) on EEG data to **classify** OCD with an accuracy of 70%
  - Collaborated on 2 research papers working with a **multi-disciplinary team** and presented the results at the 1st Sharif Neuroscience Symposium in 2018

TECHNICAL SKILLS

Python | R | Matlab | PySpark | R stan (probabilistic programming) | Deep Learning (TF, Keras) | Machine Learning | HTML | SQL | Linux | Git | AWS | GCP | Hadoop | Cassandra | Docker | Kubernetes | REST API | GraphFrames | Shiny | Big Data Engineering | Big Data Analysis | Stream Processing | Statistical Analysis | Data Modeling | NLP | Data Visualization | Data Warehousing | ETL Pipelines | Model Training & Deployment | Investment Portfolio | Investment Risk Management | Tableau | Bayesian Statistics | Flask | DSA | OOP | Power BI | Excel | Weka | Google Analytics |

SOFT SKILLS

Teamwork | Research | Critical Thinking | Problem Solving | Leadership | Time Management | Effective Communication with Stakeholders | Storytelling | Attention to Detail |

CERTIFICATES

- Python Data Structures and Algorithms | LinkedIn Learning
- AWS Fundamentals: Going Cloud-Native | Amazon Web Services
- AWS Fundamentals: Addressing Security Risk | Amazon Web Services
- Portfolio and Risk Management | University of Geneva
- Investment Risk Management | Coursera Project Network
- Visualizing Citibike Trips with Tableau | Coursera Project Network
- Natural Language Processing with Classification and Vector Spaces | DeepLearning.ai
- Modernizing Data Lakes and Data Warehouses with GCP | Google Cloud
- Google Cloud Platform Big Data and Machine Learning Fundamentals | Google Cloud