ydnt/global//global/global 1Hens2018aydnt/global//global/global 1Hens2018ydnt/global//glob
1Paixao2018ydnt/global//global/global 1Talattof2018ydnt/global//global/global 1Hens2017aydnt/
1Hens2017ydnt/global//global/global 1Chow2016ydnt/global//global/global 1Talattof2016ydnt/gl

# arjangtalattof

principal data scientist

arjang@umich.edu
linkedin
github
flickr
+44 7555117012
+01 6467700275

## languages

trilingual english/french/persian
conversational spanish & italian

## technologies

Machine Learning (Gaussian Processes, SVM, Neural Networks, Decision Trees, Random Forests)

Apache Kafka, Spark, Zookeeper

Cassandra, Graph Databases, PostgreSQL, Redis

Gremlin, Java, Python (pandas, scipy, numpy, scikit-learn), R, Scala, Tinkerpop

Unix/Linux (Centos, Debian, FreeBSD, Gentoo, RedHat), OS X, Shell scripting (bash, sed, awk)

## activities

football
photography

## INTRODUCTION

Data scientist and big data engineer in London. Python, Java & Scala developer. Spark and graph database enthusiast. Applied mathematician by training. Scholar of gastrointestinal motility variation. UK Tier 1 Exceptional Talent Visa.

## technical summary

**Computer vision machine learning pipeline.** , *production*
Deploying AWS Batch, SQS, RDS (PostgreSQL), Cassandra and Spark to handle large computer vision end-to-end processing pipeline:
ETL and data abstraction layer to allow computer vision engineers access to and from the backend without having to know specific query languages (SQL, CQL, Gremlin)
Evaluation of neural network performance regarding feature extraction for global image retrieval
Using LSH techniques to perform global image retrieval based on features extracted from neural networks.

**Massively scaled graph database analytics.** , *production*
Leveraging Cassandra and Spark (Spark Streaming) for large-scale graph networks and analyses including:
Building tools to explore and analyze graph data in a distributed cloud-based cluster.
Developing machine learning algorithms and automation of real-time entity resolution / disambiguation at scale.
Dealing with $10^7$ (eventually reaching $10^9$) transactions daily utilizing Kafka and Spark Streaming to ingest massive amounts of data through an ETL pipeline. successfully deploying working beta software to client; revenue increase from $1.8M to $6.0M; internal investment by firm ($0.5M) to generalize new capability based on client deliverable.

**Stochastic fluid transit model** , *research*
Developed in Java to to study the effects of physiological variations on drug transit, dissolution, and absorption
Pulsatile, non-deterministic approach;
Application of a non-homogeneous Poisson process;
Based on sampling from and imaging of small bowel free water content

**Analysis of noisy gastrointestinal pressure signals** , *research*
Wavelet-based peak detection and kernel density estimation; Machine learning, Gaussian process regression & PCA for signal classification; Completed in Python with SciPy/NumPy/SciKit packages

**Java interface for handling parameters and their values with syntax highlighting** , *production*
Managing dynamic parameters that control processing pipeline; Parameters and database fields (JDBC Data_Port for MySQL); Pipeline source queue management and error handling

## professional experience

aug 17 - pres **Scape Technologies** , *principal data sciencist*

*Building core date flow and analysis pipeline for location-based recognition, allowing devices to see and remember their surroundings & augment the world around them. Cloud side infrastructure allows ordinary mobile devices to enhance the world around them by overlaying digital items onto the physical world, both indoors and outdoors, using machine vision & artificial intelligence.*

aug 16 - jul 17 **Deloitte Consulting LLP** , *senior data science consultant*

*Mission Analytics in Business Model & Transformation/Strategy & Operations. Supporting senior government executives in the development of the organization's strategy and business process; assisting in development, collection, analysis, and reporting of data by leverage big data and machine learning technologies; serving as domain knowledgeable resource in advising the Deloitte team and client on tools and techniques to improve workflow.*

jul 15 - aug 16 **U.S. Food and Drug Administration** , *research fellow*

*Division of Quantitative Methods and Modeling in the Office of Research and Standards within the Office of Generic Drugs. Applying mathematical analysis to physiological/molecular based models for drug absorption, bioavailability, distribution and effectiveness. Using large data sets to improve the prediction and regulatory decision making for generic drugs.*

mar 06 - aug 07 **Department of Lunar & Planetary Sciences**
**University of Arizona** , *java developer*

*Developing open-source software for data analysis for the HiRISE instrument on the Mars Reconnaissance Orbiter. Co-funded by NASA and JPL.*

## education

2015 **Ph.D.** Pharmaceutical Sciences , *University of Michigan*
Dissertation: *Mechanistic Analysis and Quantification of Gastrointestinal Motility: Physiological Variability and Plasma Level Implications.*

2011 **M.Sc.** Pharmaceutical Sciences , *University of Michigan*

2009 **M.Sc.** Computational Biology , *New York University*

2006 **B.Sc.** Mathematics , *University of Arizona*

## publications