Kazem Jahanbakhsh

Cell: +1 (778) 990-9140 - Email: k.jahanbakhsh@gmail.com
Homepage: `http://www.kazemjahanbakhsh.com`

SUMMARY

I received my PhD in computer science with research focus on designing unsupervised and supervised machine learning algorithms to predict how people move and interact in social environments such as conferences and outdoor events.

Over the last six years, I have led a few data science teams designing and implementing ML and NLP algorithms for a range of applications including: Online Advertising, Fraud Detection, Sentiment Analysis, Text Summarization, Text Tagging, Search Ranking, and Named Entity Recognition.

Technical skills: Python, Java, C++, PhP, SQL, Apache Spark, BigQuery, AWS, scikit-learn, R.

EDUCATION

PhD, Computer Science, University of Victoria, Canada 2007 - 2012
M.Sc., Electrical Engineering, Sharif University of Technology, Iran 2003 - 2005
B.Sc., Electrical Engineering, Sharif University of Technology, Iran 1998 - 2001
.

ACADEMIC AWARDS & HONORS

Ranked *3rd* in Canada & *44th* in the world in IEEEXtreme Programming Competition. 2011
University of Victoria Graduate Scholarship ($30,000). 2007 - 2008
Ranked *50th* among 10,000 students in Iranian Graduate Studies Entrance Exam. 2003
Ranked *24th* among 300,000 participants in Iranian University Entrance Exam. 1997

WORK EXPERIENCE

Senior Data Scientist, PlentyOfFish (Match Group), Vancouver 2017 - Present

- Designing and implementing an Scam/Bot ML detection system by modeling users behavior on PoF website & mobile apps.
- Optimizing matching algos for US by modeling & incorporating users ethnicity preferences.

Data Science / Tech Lead, Qudos, Vancouver 2015 - 2017

- Led a data science team working on a tagging system composed of a number of classifiers: KNN, Decision Tree, Naive Bayes, SVM. The tagger was taking a company unstructured data as input to predict their industry tags such as Tech, Advertising, Real Estate etc.
- Supervised the design/evaluation of a Linear Regression model to improve Qudos algorithm for ranking firms. The regression model had to incorporate a number of user input filters and firm attributes such as industry tags, B2B service tags, locations, size etc.
- My other responsibilities were product design, funnel analysis, data scraping/parsing, SEO, and optimizing advertiisng campaigns.

Co-Founder, A.I. Optify, Vancouver 2014 - 2015

- Designed and implemented a customized RTB bidder for Roadhouse Interactive mobile game studio.
- Designed an ML model responsible to take a bid request from MoPub ad exchange and compute an optimal bid value for User Acquisition for Bingo mobile game.

- A Logistic Regression model was trained on a number of features (e.g. creative, time, user location, user interests, OS, app name, browser etc) to compute the conversion probability and the optimal bid for a bid request. The ML tech stack was Python, Amazon EC2, S3, & Apache Spark.

Chief Data Scientist, Trulioo, Vancouver 2013 - 2014

- Analyzed 1.2 billion Facebook & Twitter profiles to identify significant positive and negative indicators to classify authentic and fraudulent accounts.
- Engineerd a number of features for classification such as account age, user's friends graph connectivity distribution, profile/cover images, tweeting times distribution, posted contents and many more.
- Designed and tuned a Rule-based Classifier to take into account the positive and negative indicators.
- Designed and prototyped a Rule-based system to detect/stop fraudulent accounts who were using FundRazr Crowdfunding platform for money laundering.

Data Scientist, Seeker Solutions, Victoria 2013

- Implemented Latent Dirichlet Allocation (LDA) model for extracting hidden topics from medical documents.
- Feature engineering, implementing, and evaluating Hidden Markov Model (HMM) and Conditional Random Field (CRF) for tagging medical documents with rare classes such as DISEASE and TREATMENT.

Research Assistant, University of Victoria, Canada 2007 - 2012

- Designed and implemented several unsupervised and supervised learning algorithms (i.e. Random Walk, Shortest Path, Social Similarity, KNN, Logistic Regression) to predict the social interactions among people in various environments such as academic conferences and outdoor events.
- Empirical and theoretical analysis of the running time of several information spreading algorithms on the Facebook social graph.

Network Engineer, Patsa, Tehran 2006

- Worked on *Iran MPLS* project for redesigning the Internet backbone. Configuring *Sun* Servers and *Oracle* database servers. Installation & configuration of high end *Cisco* routers' management applications.

Research Assistant, Sharif University of Technology, Iran 2004 - 2005

- Designed and implemented a parallel algorithm in *C* to crack RSA-Keys using Message Passing Interface. The parallel code was running on a 17-nodes Linux Cluster where we could crack a 330-bit RSA key in less than 24 hours.

Software Engineer, Afranet, Tehran 2003 - 2005

- Implemented Sanjesh.org and Azmoon.com websites using *LAMP* technologies and Linux Clustering.
- Linux servers administration & configuration including *Apache*, *MySql*, and *Oracle* database servers.

PUBLICATIONS

- K. Jahanbakhsh and Y. Moon, The Predictive Power of Social Media: On the Predictability of U.S. Presidential Elections using Twitter, 2014, submitted to arxiv.
- K. Jahanbakhsh, V. King, G.C. Shoja, Predicting Missing Contacts in Mobile Social Networks, Elsevier Pervasive and Mobile Computing Journal, 2012.
- K. Jahanbakhsh, V. King, G.C. Shoja, Predicting Human Contacts in Mobile Social Networks using Supervised Learning, The Fourth ACM Annual Workshop on Simplifying Complex Networks for Practitioners, 2012, Lyon, France.
- K. Jahanbakhsh, V. King, G.C. Shoja, Empirical Comparison of Information Spreading Algorithms in the Presence of 1-Whiskers, Third IEEE International Conference on Social Computing, 2011, MIT, Boston, USA.
- K. Jahanbakhsh, V. King, G.C. Shoja, Predicting Missing Contacts in Mobile Social Networks, IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks, 2011, Lucca, Italy.
- K. Jahanbakhsh, G.C. Shoja, V. King, Human Contact Prediction Using Contact Graph Inference, The Third IEEE/ACM Conference on Social Computing, 2010, Hangzhou, China.
- K. Jahanbakhsh, G.C. Shoja, V. King, Social-Greedy: A Socially-Based Greedy Routing Algorithm for Delay Tolerant Networks, ACM/SIGMOBILE MobiOpp, 2010, Pisa, Italy.
- Y.O. Yazir, K. Jahanbakhsh, S. Ganti, G.C. Shoja, Y. Coady, A Low-Cost Realistic Testbed for Mobile Ad-hoc Networks, IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2009, Victoria, British Columbia.
- M. Ghelichi, K. Jahanbakhsh, E. Sanaei, RCCT: Robust Clustering with Cooperative Transmission for Energy Efficient Wireless Sensor Networks, 7th International Conference on Information Technology : New Generations, 2008, Las Vegas, Nevada, USA.
- K. Jahanbakhsh, M. Hajhosseini, Improving Performance of Cluster Based Routing Protocol using Cross-Layer Design, 2008, submitted to arxiv.
- K. Jahanbakhsh, J. Papadopoulos, An efficient Parallel Implementation of Self Initialization Quadratic Sieve for Integer Factorizations Using Message Passing Interface (MPI), Proceedings of 14th Iranian Conference on Electrical Engineering, 2006, Tehran, Iran.

SOFTWARE RESEARCH PROJECTS

Machine Learning/Data Mining/NLP projects 2007 - Present

- *Predicting US 2012 Election*: a research project to predict US 2012 presidential election through sentimet analysis of a large number of tweets. This work was covered on Forbes.
- *Information Spreading*: an efficient C code for empirical analysis of information spreading algorithms on Facebook social graph.
- *Social-Sim*: a simulator written in *C++* for statistical analysis of mobile social graphs.
- *Human Contact Predictor*: a Python implmentation of several unsupervised learning algorithms for predicting how people interact in social settings such as conferences.
- *Hometown Predictor*: an algorithm implemented in *Python* for predicting where a person lives by analyzing her Flickr photos. This work was covered on MIT Technology Review.
- *K-means Clustering*: a *Python* implementation of *K-means* clustering algorithm.

- *Community Detection*: a *Python* implementation of Girvan-Newman *community detection* algorithm for weighted social graphs.
- *Flickr Crawler*: a two-layer crawler in *Python* for collecting *Flickr* photos using Flickr API. The first layer crawls Flickr social graph while the second layer crawls Flickr users profiles and their uploaded photos tags.
- *Reliable Datagram Protocol*: a multi-threaded reliable application layer implemented in *C*. This application layer runs on top of UDP to make it reliable.
- *Soma Cube Puzzle*: *Java* code for solving 7-pieces Soma Cube puzzle by using a backtracking search.
- *Flying Blimp*: an embedded system developed in *C* for controlling an autonomous flying blimp.

COURSES TAKEN

Data Mining (A+), Algorithmic Mechanism Design and Social Computing (A+), Randomized Algorithms (A+), Analysis of Algorithms (A+), Topics in Artificial Intelligence (A), Software for Embedded & Mechatronics Systems (A+), Wireless & Mobile Networks (A-), Communication Networks (A+), and Operations Research & Simulation (A+)

Activities

I enjoy playing soccer, chess, and cooking.