JUDE KHOUJA   +1 202 733 0000   mkhooja@gmail.com

SKILLS
NLP: MC, QA, LM, Language Representation **Deep Learning:** RNNs, Attention, Seq2Seq
Supervised ML: Linear Models, SVM, Random Forests. **Unsupervised ML:** PCA, ICA, Hierarchical Clustering, LDA.
DL frameworks: Pytorch, Tensorflow **Programing** Python, SQL

EDUCATION
Stanford University - Graduate coursework in Natural Language Processing with Deep Learning, 2017
University of Maryland College Park - Masters in Information Management, May 2012
Damascus University - B.E. Computer Science, Focus: Artificial Intelligence, July 2007

HONORS   **Fulbright** Scholarship, U.S. Department of State 2010.   ## LANGUAGES   Arabic: Native   English: Fluent

HIGHLIGHTED EXPERIENCE

May 2018 – Sep 2019
SENIOR APPLIED SCIENTIST, MICROSOFT, CA
Researching and building large-scale Language Models for various speech services at Microsoft.

May 2016 – Feb 2018
PRINCIPAL DATA SCIENTIST, SALESFORCE, CA
Built text classification, text representation internal tools using deep learning methods (FFN, RNNs) to automate several customer support processes.

September 2015 – April 2016
SENIOR DATA SCIENTIST, SALESFORCE, CA
Built models for churn prediction and customer segmentation using various classification, regression and clustering methods.
Developed reusable machine learning pipeline and feature engineering libraries.

Jul 2014 – Nov 2014
DATA SCIENCE LEAD, CRITTERCISM, CA
Applied clustering techniques and text matching for cardinality reduction.
Built internet facing analytics portal for tracking mobile performance metrics worldwide.
Implemented AWS on demand analytics infrastructure using S3 and EMR.

Nov 2013 – Jun 2014
DATA ANALYTICS LEAD, IREX, DC
Developed and implemented data driven practices and programs for violations documentation.
Managed and oversaw the building of the technology and the data team from scratch.

Mar 2013 – Oct 2013
BIG DATA SPECIALIST, ORACLE, VA
Developed machine learning proof of concepts in the public sector using technologies including Hadoop, Hive, Pig, R Enterprise, Mahout and other proprietary and open source tools.

Jun 2012 – Aug 2012
DATA SCIENTIST & BIG DATA ENGINEER, ORACLE, VA
Was the Lead Data Scientist in the Oracle/NCI partnership project which resulted in winning the "2012 Best Government Big Data Solution" Award.

Developed MapReduce programs in Java and Python for analyzing simulated Gene data of 900 million patients.
Assisted in building a 9-node Hadoop (CDH) cluster.

Oct 2011 – May 2012
GRADUATE RESEARCH ASSISTANT, UMIACS, MD
Evaluated large scalable Topic Modeling algorithms (LDA) based on distributed frameworks.
Conducted lexicon expansion using LDA algorithms.
Developed an Arabic version of the Word Count tool for sentiment analysis and honor dictionary validation.

Aug 2009 – Aug 2010
PROGRAM OFFICER – M&E, UNITED NATIONS, SYRIA
Developed and implemented a comprehensive M&E plan for a $6 Million grant by The Global Fund Program (GFATM).
Oversaw the design and implementation of the health management information system.

Aug 2008 - Mar 2009
TECHNOLOGY CONSULTANT, UNITED NATIONS, SYRIA
Conducted comprehensive assessment of the existing IT infrastructure and capacity for the United Nations (UNDP) project on Aid Effectiveness and produced a full analysis report to the Minister of state planning commission.

TRAINING
M&E for UN programs, United Nations, Geneva 2009.
Shell LiveWire Entrepreneurship Program, Damascus 2009.
SCJP (Sun Certified Java Programmer), Sun 2008.

PUBLICATIONS
Knowledge and Theme Discovery across Very Large Biological Data Sets Using Distributed Queries: A Prototype Combining Unstructured and Structured Data. PLOS 2013

Mr. LDA: A Flexible Large Scale Topic Modeling Package using Variational Inference in MapReduce. ACM International Conference on World Wide Web, 2012.

Social Media for Political Change: The Activists, Governments and Firms Triangle of Powers During the Arab Movement. Handbook of Research on Political Activism in the Information Age (pages 26-36). IGI Global 2015