Sheikh Shams Azam

azam1@purdue.edu
https://www.linkedin.com/in/shams-sam/ +1-(765)775-9420

Education

- **Purdue University** — West Lafayette, IN
  Ph.D in Electrical and Computer Engineering; GPA: 4.0/4.0 — *August 2019 – Present*
- **National Institute of Technology Karnataka (NITK)** — Surathkal, India
  B.Tech in Electrical and Electronics Engineering; GPA: 8.28/10.0 — *July 2011 – May 2015*

Publications

- **Exclusion-Inclusion Generative Adversarial Nets (EIGANs)**
  Grant under Northrup Grumman Cybersecurity Research Consortium; Under Review, ICML 2020
  - **Abstract**: A novel GAN architecture consisting of three types of networks: (i) an encoder network, (ii) m adversary networks, and (iii) n ally networks, with the objective of learning data encodings that are private yet informative, i.e., predictive of intended targets while obfuscating sensitive attributes in the original dataset
- **CASCADENET: A Hierarchical LSTM Architecture for Automated ICD-10 Coding**
  Funded by Foundation AI; FICC-2019
  - **Abstract**: A novel hierarchical neural network architecture (hybrid of Dense and LSTM layers) for accurate classification of medical diagnosis text among massively categorical ICD-10 codes (approx. 93k classes) using 100k training examples incorporating distributed word embedding (Word2Vec)
- **Q-Map: Medical Concept Mining from Clinical Documents**
  Funded by Practo Technologies; ICSWNLP-2018
  - **Abstract**: A medical concept mining system running UMLS Metathesaurus concept search on clinical free text inputs that employes Aho-Corasick finite state machine and detects context based negation. Q-Map overcomes shortcomings of its contemporary MetaMap released by National Library of Medicine (NLM)

Research Experience

- **Purdue University** — West Lafayette, IN
  Graduate Research Assistant — *August 2019 - Present*
  - **Privacy Preserving Distributed ML Architectures**: Development of a multi-agent conditional GAN architecture with the objective of training an encoder to learn data representations that are private yet informative and it's extension to decentralized training for application in real-word scenario of sensor data security deployed in distributed network field
  - **Link Prediction in Large Scale Social Networks**: Prediction of links in large scale networks using network characterizations derived using unsupervised classical network centrality/closeness metrics as well as supervised deep neural networks using Graph convolutions, and it's real work application in social learning networks for improving overall learning experience
- **Foundation AI** — Los Angeles, CA
  Research Scientist — *September 2018 - August 2019*
  - **Document Analysis and Optical Character Recognition**: OCR pipeline for privacy-oriented offline deployments: image processing such as, orientation and alignment correction by characterizing hough transform of processed images and background/noise removal by applying adaptive/median filtering, scene text detection using EAST and Mask R-CNN, binarisation using cGANs, character recognition using CNNs.
  - **Secure Containerization**: Standardization of deployment and development platform using containerization and distribution of tasks over services running on containerized micro-service architecture with API interfaces and secure communications by a docker SSL layer (dynamically manages updates in SSL certificates)
- **Practo** — Bangalore, India
  Data Scientist, Senior Software Engineer, Software Engineer — *June 2015 - August 2018*
  - **Concept Mining and Automated Coding from Clinical Documents**: Developed methods such as Q-Map and CASCADENET for automated coding ICD-10 (International Code for Diseases), CPT (Current Procedural Terminology) of clinical documents, which plays a vital role in subsequent tools including but not limited to insurance claims, disease trend and epidemic outbreak characterization

- **Practo Search**: Scalable system for faster search and intelligent suggestions reliant on data driven adaptive ranking algorithm for personalized search results. Extending system support to allow multi-language and accent-agnostic search using language modelling and knowledge source graphs
- **Text Preprocessing and Adaptive Stopwords Filtering**: NLP based text preprocessing for noisy channels such as HTML, RTF etc. and stop-word removal by thresholding Kullback Leibler Divergence and Inverse Document Frequency
- **Practo Discovery**: Web server level Url Discovery and Redirection System for preventing web traffic overload on Application Servers and protection against DDoS attacks

- **Simple Wealth** Bangalore, India
  Machine Learning Intern *December 2015*
  - **Smart Offers Recommendation**: Fetching media, by implementing scraping tools for websites and classification for personalized offer recommendations based on popularity, user interests and geolocation by leveraging concepts of logistic regression, support vector machines and decision trees

Undergraduate Projects

- **Application of Machine Learning in Image Segmentation** Advisor: Dr. Ashvini Chaturvedi
  Major Project *July 2014 - May 2015*
  - **Abstract**: Analysis of algorithms (k-mean clustering and morphological edge detection) for isolation of object boundaries images and enhancing techniques to perform better on noisy images by employing classical image processing techniques supported by statistical estimations
- **Application of Object Tracking Algorithm in a Video** Advisor: Dr. Ashvini Chaturvedi
  Mini Project *July 2013 - December 2013*
  - **Abstract**: Conception and implementation of algorithm for tracking trajectory of an object in a video using image processing techniques such as frequency domain analysis and 2D-DFT
- **Speech Recognition System** Advisor: Dr. K Manjunatha Sharma
  Mini Project *July 2013 - December 2013*
  - **Abstract**: Basic speech recognition system using total energy algorithm and zero-crossing algorithm to control speed operation of AC motor

Competitions and Projects

- **Machine Learning Medium**: Author of Machine Learning Medium (https://machinelearningmedium.com), an educational website, deployed and maintained on github pages using Jekyll and hosted on CloudFlare
- **Kaggle: Data Science Bowl 2017**: Among top 13% in Kaggle Data Science Bowl, 2017. Developed a machine learning classifier based on ResNet using 3D convolution for classification of malignancy of lung nodules from 3D CT Scans
- **Kaggle: Quora Questions Pairs**: Among top 16% in Kaggle Quora Question Pairs Challenge. Developed a LSTM, XGBoost transfer learning model for pairing of similar questions on Quora, a QnA website
- **HackerEarth, IndiaHacks 2017**: Secured rank 18 among several thousand participants at IndiaHacks Challenge, 2017 organized by HackerEarth for solving real world classification problems of HERE maps and Hotstar

Awards And Achievements

- **Young Leader, ISB**: One of two students selected for ISB Young Leaders Program (ISB-YLP) in Senior Year during B.Tech
- **Finisar Malaysia Funded Internship**: One of three students selected for Finisar Malaysia Funded Internship during Junior year
- **Highest Campus Placement, Electrical Engineering, NITK Class of 2015**: Highest campus placement among students of my major subject (class of 110)
- **Academic Scholarships**: Selected to receive various Scholarships for Academic Excellence for Undergraduate and High-school Studies including Indian Air Force-BA Scholarship and KVS Scholarship for AISSE and AISSCE Certificate CBSE Examinations

*References available on request.