

# SOUVIK DUTTA

509 W Main St, Apt 8, Urbana IL 61801, USA

[sdutta9@illinois.edu](mailto:sdutta9@illinois.edu), +1 (217) 735-0789

GitHub: [/TheKivs](#) LinkedIn: [/souvik-dutta](#)

- Ph.D. candidate with experience in Machine Learning algorithms and Big Data with **Google Cloud, Spark & AWS\***
- 3+ years academic experience with ML libraries like **Scikit-learn, Pandas, TensorFlow (PyTorch & Keras wrappers)**
- Research interest in GANs, NLP, VAEs, Graphical Models, Computer Vision: specifically, at using generative and discriminative models for language-vision tasks \*AWS Certified Cloud Practitioner

## EDUCATION

- **Ph.D. candidate**, University of Illinois Urbana-Champaign, USA August 2020
  - ◊ **Field of research**: Mathematical Physics, Advisor: Dr. Thomas Faulkner [Publications]
  - ◊ **Relevant Courses**: Machine Learning, Data Mining, Deep Learning, Computer Vision
- **B.Tech.**, Indian Institute of Technology Bombay, India May 2013
  - ◊ **Major**: Engineering Physics (with Honors)
  - ◊ **Relevant Courses**: Data Structures & Algorithms, Optimization, Data Analysis & Interpretation, Linear Algebra
  - ◊ **Thesis**: "Efficient clustering algorithms at particle colliders", received "Undergraduate Research Award"

## WORK EXPERIENCE

- **Graduate Research Fellow**, University of Illinois Urbana-Champaign, USA Aug 2015 - present
  - ◊ Applied **optimization techniques** to improve noise-resilience in quantum computing circuits by **~20%**
  - ◊ Devised **2 novel algorithms** for solving high-dimensional complex optimization problems in polynomial time [talk]
- **Teaching Assistant**, University of Illinois Urbana-Champaign, USA Sep 2013 - Jul 2017
  - ◊ Designed **R/Python3** tutorial sessions for a class of 90 graduate students on Statistical Data Analysis [link]
  - ◊ Instructed **coding sessions** for courses on Nonlinear Optimization, Quantum Computation, and Fluid Dynamics
  - ◊ Received the prestigious Illinois Chancellor's "Outstanding Teaching Assistant" award 4 times
- **Quantum Computing Research Intern**, University of Milan, Italy May 2013 - Jul 2013
  - ◊ Conducted numerical optimization in 16-qubit lattices to simulate fault-tolerant **data encryption** protocols
  - ◊ Led 3-tier team of **24 interns** to design, simulate and A/B test memory-efficient quantum key-distribution algorithms
- **Machine Learning Research Intern**, CERN, Switzerland May 2012 - Aug 2012
  - ◊ Spearheaded the large-dataset preprocessing and **hierarchical clustering** library "FastJet" in Python [paper]
  - ◊ Deployed algorithm into ML pipeline; resulted in  $\mathcal{O}(N^2)$  speedup and **~\$1.2M/FY** savings in grid-time expenses

## ACADEMIC PROJECTS

- **Edge detection for 3D brain MRI reconstruction**, University of Mainz, Germany [link]
  - ◊ Implemented scalable filtering techniques in **computer vision** using **BigQuery** to detect edges in low-res MRI scans
  - ◊ Achieved image segmentation and tumor detection with **77% accuracy** using **TensorFlow with CUDA**, Scikit-learn
- **Image Captioning Conditioned on Part-of-Speech**, University of Illinois Urbana-Champaign, USA
  - ◊ Obtained state-of-the-art accuracy (best-1,  $k^{th}$ ) at **caption generation** over existing GAN and VAE-based methods
  - ◊ Achieved **high diversity** evaluated by distinct  $n$ -grams, novel sentences and mBleu-4 (overlap) scores as metrics
- **Churn detection and intervention**, FlipKart Co., India
  - ◊ Trained and deployed a **16%** more efficient end-to-end classification pipeline on **AWS EC2 (g2.2xlarge) GPUs**
  - ◊ Predicted existing customer churning with **84% accuracy** (AUC 0.87), with potential profits of **\$160,000/FY**

## TECHNICAL SKILLS

- **Programming**: Python, R, C++, Java, SQL (Postgre, BigQuery), Tableau, MATLAB, GNU Octave
- **Libraries**: TensorFlow, PyTorch, Keras, XGBoost, Scikit-learn, Pandas, NumPy, SciPy, Matplotlib, Seaborn
- **Algorithms**: Bayesian classification, Linear & Logistic regression, KNN, K-means clustering, Decision Trees, SVM, Ensemble learning, Reinforcement learning, Regularization theory, Hidden Markov Models, CNNs, Q-learning
- **Mathematics**: Linear algebra, Probability theory, Multivariate vector calculus, Optimization, Graph Theory