Curriculum Vitae
PhD, School of Information

AGRIMA SETH

4007 Chestnut St., Philadelphia, PA 19104 | agrima@seas.upenn.edu | +1 (682) 202-2742 | www.linkedin.com/in/agrimaseth
https://github.com/agrimaseth | http://agrimaseth.github.io/

EDUCATION

University of Pennsylvania, School of Engineering and Applied Science, Philadelphia, PA — **May 2018**
Candidate for Master of Science in Engineering, Major: Computer & Information Science — GPA: 3.88/4.00

University of Pune, India — **May 2016**
Bachelor of Engineering, Information Technology — GPA: 80/100

RELEVANT COURSEWORK

Machine Learning | Computational Linguistics | Database and Information Systems | Information Storage & Retrieval
Analysis of Algorithms | Software Systems | Computer Vision & Computational Photography | Data Structure

PUBLICATIONS & SCHOLASTIC ACHIEVEMENTS

- Seth A., Nayak S., Mothe J. and Jadhay S. (2017). News Dissemination on Twitter and Conventional News Channels In Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS, ISBN 978-989-758-247-9, pages 43-52. DOI: 10.5220/0006264100430052.
- Agrima Seth, Dr. Deepak Mishra. Comparative Study of Geometric and Image Based Modelling and Rendering Techniques, arxiv.org/abs/1409.5024 Citation: arXiv: 1409.5024.
- Peer reviewed book titled "BeagleBone Black Cookbook" (ISBN 13 9781783982929).

TECHNICAL PROFICIENCIES

Languages: C, Python, C++, Java, LISP, JavaScript, Java Servlets, PHP, node.js, Matlab, R.
Operating Systems: Windows, Unix, Linux.
Libraries and Packages: NLTK, scikit-learn, Numpy, Scipy, Pandas, Opengl, OpenCV, XML, JSON, Protobuf, Keras, Tensorflow.
Databases & Tools: Git, NoSQL, MySQL, Cassandra, DynamoDb, Weka, 3d Max Studio, Blender, Vizard 3D, V-ray Renderer, Tableau.
Knowledge of natural language processing, machine learning, feature engineering, data mining, data analysis and data visualization.

EXPERIENCE

Master's Thesis, University of Pennsylvania — **Philadelphia, PA (Jan 2017 – Dec. 2017)**

- Performed a study on "Depression Analysis Using Facebook and Twitter Posts" to develop a robust cross platform model.
- Engineered features like LIWC, LabMT and Topics to identify the lexica of a depressed individual. Currently, analyzing effect of age, gender and demography on lexical expressions.
- Identified the most robust features and analyzed the performance of predictive models to develop cross platform model.

Machine Learning Intern, European Organization for Nuclear Research (CERN) — **Geneva, Switzerland (June 2017 – August 2017)**

- Collaborated with Compact Muon Solenoid (CMS) EP-CMG team at CERN & implemented normalization techniques in **Python** to study data patterns in drift tubes.
- Developed machine learning test model (autoencoder) using **Keras & Tensorflow** to automate current paradigm of quality assessment by detector experts facilitating checking of large volumes of data in real-time improving ability to detect unexpected anomalies.
- Served on core machine learning research team, experimented with different machine learning models and feature selection techniques. **Awarded 2nd prize** among 37 intern projects during Openlab Lightning talk at CERN, Geneva.

Software Developer (Penn Cloud), University of Pennsylvania — **Philadelphia, PA (March 2017 – April 2017)**

- Designed and implemented distributed cloud service using protobuf and plain text for communication. Created fault-tolerant backend server using quorum based replication that employs key-value store similar to Google's Bigtable using **C++.**
- Built HTTP server to process request from users by retrieving appropriate data from backend server and constructing HTTP response for GUI. Handled SMTP, storage services and persistent HTTP connection.

Research Intern, Institut de Recherche en Informatique de Toulouse — **Toulouse, France (Dec 2015 – Feb 2016)**

- Deployed **hbase** to store Twitter data. Drew comparative results on flow of catastrophic topics on Twitter and news channels using **Python programming**. Performed visualization on **Tableau**.
- Published paper at **ICEIS 2017 (Portugal)**. **Awarded 1st position** in Amalgam, alumni sponsored competition, AIT Pune, 2016.

Research Intern, Institut de Recherche en Informatique de Toulouse **Toulouse, France (May 2015 – July 2015)**

- Indexed collection of about 50 million web pages using Terrier and evaluated using ndeval.
- Evaluated by chunks against the entire collection as a unit, showed an improvement in effectiveness and efficiency parameters.
- Accomplished automatic expansion of a query based on relevant information using graph theory using python programming

Malaviya National Institute of Technology (MNIT) **Jaipur, India (Nov 2014 – Dec 2014)**

- Tracked Hadoop Cluster's health using Ganglia Tool.
- Implemented Single Label SVM for classifying the logs. Successfully re-created the methods proposed in paper "Fault Management in Map-Reduce through Early Detection of Anomalous Nodes" by Selvi Kadirvel, Je rey Ho, Jos A. B. Fortes

PROJECTS AND COMPETITIONS

- Created a **text summarization and analytics platform** that works on top of messaging systems being used at CERN, Switzerland. Programmed various **natural language processing and machine learning routines** for answering users' queries (summary, most important chats and keywords). Link: https://github.com/parityapp/ (Team of 2, July 2017).
- **Classified text data in various levels of difficulty** using labeled excerpts to train models using **Python, Scikit-learn**. Engineered syntactic and semantic features for model training and compared Spearman Correlation of these models on held-out dataset (March 2017 - April 2017).
- **Classified tweets as joy/sad** utilizing 4,500 labeled & 4,500 unlabeled training examples to train models. Compared Logistic Regression, Naïve Bayes and K-means for classification using **Matlab**. Improved efficiency using Word2vec, n-gram and auto-encoder features (Nov 2016 - Dec 2016).
- Employed **dlib** for **face detection** and land marking the images in each test video frame and replacement image for **automated face replacement**. Self-coded Delaunay triangulation based on land marking. Employed **OpenCV** for warping and applied Poisson blending (Dec. 2016).
- Implemented Canny Edge detection for a RGB image using Gaussian Smoothing, Non-max suppression and hysteresis to link edges based on high and low magnitude thresholds (September 2016).
- Implemented **Image Morphing** using **Triangulation** and **Thin Plate Spline** method (October 2016).
- Performed **Panorama creation** using **RANSAC** to get a consistent homography (October 2016).
- Created an online quiz with data (csv files, images) stored on **Mlab** (for Mongodb) & **RDS** (for SQL). Implemented User-Interface using **PHP**. Implemented Query-Optimization using views and indices for making dynamic question generation faster.

SCHOLASTIC ACHIEVEMENTS

- Ranked 2nd among 4789 students of Information Technology in all colleges of University of Pune.
- Selected for MIT Media Lab 5thDesign Innovation Workshop (200 students selected out of 1500 applicants).
- Awarded Tata Merit Scholarship and Academic Merit Scholarships for standing 1st in Information Technology Dept. 2012-2016

MENTORING

- Mentored a student during the Learn,IT Girl ! program to design and code a Korean Music recommendation System. October 2017-Jan 2018.
- Cofounded Big Data Club at Army Institute of Technology, Pune India in 2015.