Amey Agrawal
agrawalamey.github.io

Email: agrawalamey12@gmail.com
Github: /AgrawalAmey
+91-9928852500

Education

- **Birla Institute of Technology and Science** — Pilani, India
  Bachelor of Engineering in Computer Science; GPA: 8.13 / 10.0 — *Aug. 2014 – July. 2018*

Publications

- **Learning Digital Circuits: A Journey Through Weight Invariant Self-Pruning Neural Networks**
  Amey Agrawal, *and Rohit Karlupia*
  Proceedings of New in ML Workshop, NeurIPS, 2019, Vancouver [Paper][Code]
- **Delog: A Privacy Preserving Log Filtering Framework for Online Compute Platforms**
  Amey Agrawal, *Abhishek Dixit, Namrata Shettar, Darshil Kapadia,*
  Rohit Karlupia, Vikram Agrawal, and Rajat Gupta
  Proceedings of IEEE International Conference on Big Data, 2019, Los Angeles [Paper]
- **Logan: A Distributed Online Log Parser**
  Amey Agrawal, *Rajat Gupta, and Rohit Karlupia*
  Proceedings of IEEE International Conference on Data Engineering (ICDE), 2019, Macau [Paper]

Experience

- **Qubole** — Bangalore, India
  Member of Technical Staff-II — *Jul 2018 - Present*
  - **Qubole MLflow**: Designed and built modules end-to-end across the stack to integrate MLflow in Qubole ecosystem. Designed communication interfaces to glue the MLflow service with Qubole's Jupyterlab platform.
  - **Managed RStudio Service [Demo] [Blog]**: Integrated RStudio as a manged service within Qubole's data science platform. Contributed to **S3FS** improve cache performance and added support for persistent ACLs. Received an **early promotion** as a result of this effort.
  - **Logan [Blog]**: Developed a distributed log parsing algorithm that provides **39x** speed-up over the previous state-of-the-art. The system is used to perform real-time anomaly detection for all Apache Spark by processing over 250 million log lines every day. Received **Spotlight** award for this effort.
  - **Application Proxy**: Redesigned the application proxy layer which reduced the page load times for Apache Spark UI by over **5x**.
- **BITS Pilani** — Pilani, India
  Teaching and Research Assistant — *Dec 2017 - May 2018*
  - **Teaching Assistant [GitHub]** : Served as teaching assistant for courses on Neural Networks and Machine Learning. Design Python exercises which have been to used by more than thousand students so far.
- **Qubole** — Bangalore, India
  Software Engineering Intern — *Jul 2017 - Dec 2017*
  - **Deep Learning Cluster**: Developed beta version of Qubole's enterprise distributed deep learning platform. Received **Spotlight** award for cross-team collaboration.
- **Norah.ai** — Bangalore, India
  Machine Learning Research Intern — *May 2017 - Jun 2017*
  - **Text to Humanoid Animation**: Developed a sequence to sequence model for text to animation conversion using Keras.

Selected Projects

- **Callisto** — *Jan 2018*
  Guide: Prof. Surekha Bhanot — [Blog] [Demo] [GitHub]
  - Developed cross-platform desktop application built with Electron and Express for neural networks course.
  - Automatically sets up uniform scientific python development environment independent of operating system.
  - Allows students to download and launch assignment Jupyter notebooks with a single click.

- Evaluation mode enables programming contests using nbgrader.

- **Disentanglement learning for iris image indexing** *Jan 2018 - May 2018*
  Guide: Prof. Kamlesh Tiwari [Blog] [Google Colab]
  - Designed an autoencoder architecture to learn horizontal translation-invariant representations of normalized iris images.
  - Established proof of concept on MNIST, Fashion-MNIST and CIFAR10 datasets.

- **Deep Q-learning for autonomous warehouse robots** *Jan 2017 - Apr 2017*
  Guide: Prof. Surekha Bhanot [GitHub]
  - Implemented a Deep Q-learning algorithm to make warehouse robots which can learn to navigate autonomously.
  - Developed 2D simulations using pybox2D and 3D simulations using V-rep.
  - Implemented the neural network using keras.

- **CNN visualization toolkit** *Apr 2017*
  Guide: Prof. Sundaresan Raman [GitHub]
  - Integrated a collection of popular CNN visualization techniques into a single framework which can take any Keras CNN model as input.
  - Developed backend web server using Flask.

- **Automated news-in-shorts** *Nov 2016*
  Guide: Prof. Poonam Goyal [GitHub]
  - Latest posts from RSS feeds of multiple news agencies are aggregated to automatically generate abstracts for top stories.
  - Trending topics on twitter are mapped to clusters of news articles to identify trending news.
  - All the articles on a given trending topic are summarized using extractive text summarization using sumy.
  - Developed web clients and REST API using MongoDB and Express.
  - Implemented text processing pipeline using NLTK, Scikit-learn and Gensim.

- **Pokemon MMORPG** *Mar 2017 - Apr 2017*
  Guide: Prof. Rahul Banerjee [GitHub]
  - Designed and developed fully distributed multi-player online game entirely in C.
  - Used Redis as message broker and MongoDB as database.

- **Arxiv-Sanity v2.0** *Feb 2017*
  Open Source Initiative [GitHub]
  - Revamped Andrej Karpathy's Arxiv-Sanity for improved functionality and a better UI.
  - Built upon the existing flask web server and migrated database to MongoDB for scalability.

- **Predicting Election Results using Twitter** *Apr 2015*
  Guide: Prof. Vandana Agrawal
  - Built a CNN model with word embeddings to perform sentiment analysis on tweets pertaining to US presidential elections.
  - Implemented CNNs from the ground up in vanilla python.

Technical Skills

- **Languages**: Proficient in Python, Scala, JavaScript, Bash, Familiar with Java, Ruby, R, C/C++
- **Data Science**: PyTorch, Keras, Spark, TensorFlow, Scikit-learn
- **Web & UI**: Express, Flask, Jetty, Ruby on Rails, Electron, JQuery, React
- **Databases**: MySQL, Redis, MongoDB, SparkSQL
- **Misc**: Git, Markdown, LaTeX, Jenkins, CorelDraw

Selected Coursework

- **Mathematics**: Multivariable Calculus, Linear Algebra, Probability & Statistics, Differential Equations
- **Software Engineering**: Object Oriented Programming, Computer Networks
- **Systems**: Database Management Systems, Database Systems
- **Data Science**: Machine Learning, Information Retrieval, Neural Networks and Fuzzy Logic, Cognitive Computing