# Anmol Jagetia

ajagetia@cs.cmu.edu ❖ anmoljagetia.me ❖ (412) 708 4554 ❖ github.com/anmoljagetia

## EDUCATION

**Carnegie Mellon University, School of Computer Science** **Dec 2020**
*Masters of Computational Data Science (GPA: 3.85/4)* *Pittsburgh, PA*

- *Coursework:* Deep Learning, Machine Learning for Large Datasets, Neural Networks for NLP, Multimodal Learning, Reinforcement Learning, Computer Systems, Machine Learning, Cloud Computing, Interactive Data Science, Data Science Seminar, Capstone Project.
- *Projects:*
  - Capstone project on improving training time for DL models by leveraging compression (PCR) and lower fidelity training for audio.
  - Implemented a phoneme detector for Mel-spectrogram frames, using a 6-layer Bi-LSTM and beam-search.
  - Implemented end-to-end speech to text transcription using a p-BLSTM encoder and attention decoder with teacher forcing.
  - Machine translation and paraphrasing using a Transformer Model to output translated paraphrases achieving SOA performance.
  - Large Scale Twitter Analytics Web Service (ETL, 35K RPS web-service using AWS/Vertx/MySQL, stood 1st amongst 35 teams).
  - Implemented a taxi-fare prediction service using GCP APIs and Apache Spark.
  - Implemented malloc, a Unix shell and a multi-threaded proxy server in C.

**Indian Institute of Information Technology, Allahabad** **July 2016**
*Bachelors of Technology in Information Technology (GPA: 8.92/10)* *Allahabad, India*

## WORK EXPERIENCE

**Carnegie Mellon University, School of Computer Science** **Dec 2019 – Present**
*Teaching Assistant (Cloud Computing and Machine Learning for Large Datasets)* *Pittsburgh, PA*

- Designed, maintained and ran projects on Map-Reduce/Hadoop. The goal of the project is to introduce students to Hadoop and highlight its advantages over sequential processing. QA tested projects on ML on the Cloud, Spark, and Multi-threaded programming.
- Designed and maintained course infrastructure to offer assignments based on PySpark on Gradescope and Databricks for about 200 students. Held weekly office hours and lead recitations on Hadoop and Spark.

**Appian Corporation** **Jun 2020 – Aug 2020**
*Cloud Software Engineer Intern* *Tysons, VA*

- Developed feature adoption analytics tracker using Java, MySQL, JDBC, Appian and SAIL used for data-driven decision making.
- Developed a distributed log-aggregator using the TICK stack and Rsyslog.

**Media.Net** **Sep 2016 – July 2019**
*Operations and Data Engineer II* *Mumbai, India*

- Built and maintained cloud infrastructure on AWS and GCP using a combination of self-hosted and managed services.
- Used various Machine learning and Deep learning techniques to find anomalies and correlations in time-series data to deliver business insights for real-time bidding, ad targeting (and retargeting) for contextual advertisement and operation performance.
- Developed an AWS Sagemaker-like tool for running and maintaining distributed ML jobs on cloud infrastructure.
- Built a containers-enabled distributed cache that can scale to a few terabytes using Varnish, DDNS and Consistent Hashing.

**HPCC Systems** **May 2015 – Sep 2015**
*Google Summer of Code Student and Open Source Developer* *Allahabad, India*

- Contributed to the expansion of the visualization framework for fast, high performing cluster computing.
- Created new widgets and testing framework that works with big data within HPCC platform.

## RESEARCH EXPERIENCE

**Max Planck Institute for Software Systems** **Jun 2016 – Sep 2016**
*Research Intern* *Kaiserslautern, Germany*

- Worked with Prof. Paul Francis on anonymizing large-scale datasets with sensitive data for Machine Learning using Aircloak.
- Designed a testing suite to maintain integrity of the system using Aircloak, Python and MySQL Events.

**Complutense University (Datalab)** **Feb 2016 – Jun 2016**
*Research Assistant* *Madrid, Spain*

- Collaborated with Prof. David Gómez-Ullate and Dr. Pablo Suarez on supervised learning methods using Bayesian Statistics and random forests-based ML classifier to detect fraud in credit card transactions (Research grant by BBVA Foundation).

## CONFERENCES & PUBLICATIONS

- Invited speaker at **O'Reilly AI Conference, London (2019)** discussing object detection techniques using Transfer Learning.
- Invited speaker at **O'Reilly AI Conference, New York (2018)** discussing game-bots using Reinforcement Learning.
- Co-Authored and published the paper titled *Combining Keystroke dynamics and face recognition for user verification* at the **18th International Conference on Computational Science and Engineering, Porto, Portugal (2015)**.

## SKILLS & INTERESTS

- **Languages:** Python, Java, C, Bash (Shell), Latex, C++, Scala
- **Frameworks:** PyTorch, Tensorflow, scikit-learn, NumPy, pandas, SciPy, matplotlib, PySpark, Vertx, Grok
- **Technologies:** AWS, GCP, Azure, Microservices, Kafka, Varnish, Redis, ELK, TICK, DC/OS, Kubernetes, Docker, Terraform