Neil Sengupta

9450 Gilman Dr. La Jolla, CA | neilsgp.github.io | (858)-361-5569 | nesengup@ucsd.edu

Skills

- Experience in Python, Java, Data Analysis, Pytorch and open-source distributed systems.
- Familiar with Big Data Stack like Hadoop, Spark, SQL/NoSQL databases, Airflow and Kafka.

Work Experience

Uber, Software Engineering Intern **Summer 2018**

- Built a **Spark** Pipeline in **Scala** and **HQL** to perform data ETL from internal tools to audit safety incidents on the platform.
- Automated Pipeline run using internal **Python** Airflow System and wrote end-to-end HQL tests for data validity.
- Hardened production pipelines by introducing intermediate data quality checks in Spark to detect anomalous changes and increase confidence in production data for business analysis.
- Defined and generated new metrics for Grafana Dashboards for pipeline monitoring and alerting.

Inmar, Data Engineering Intern **Summer 2017**

- Built a **Flask-MongoDB** data import API for analytics platform and deployed microservice using **Docker**.
- Integrated streaming service on client side for parsing and uploading large data sets using **JavaScript.**
- Worked on internal tool to build a collapsible tree visualization from raw CSV datasets using **Python** and **D3**.

San Diego Supercomputer Center, Software Engineering Intern **Summer 2016**

- Implemented algorithms for comparison and high performance clustering of 3D biomolecular structures using Bio-**Java** library.
- Analyzed the usage of **Apache Parquet** for data storage.
- Benchmarked approaches for hosting compressed biomolecular structure representations.

UCSD CSE Department, Section Leader **Fall 2015 – Winter 2017**

- **Lead** and **mentored** a section of 15 students and assisted 200 students on programming skills and programming assignments for **Introduction to Computer Science: Java**.
- Helped students analyze programming logic and develop debugging skills.

Publications

- Kang, Hyeonsu B., Amoako, Gabriel, **Sengupta, Neil**, AND Dow, Steven P. "Paragon: An Online Gallery for Enhancing Design Feedback with Visual Examples". Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18. Montréal, QC, Canada: ACM, 2018, 606:1–606:13.

Selected Projects

Photographic Memory

- Full stack web application that uses WebRTC and MyoJS API to capture video feed and arm-EMG data of Alzheimer patients to tag physical actions and provide information to caregivers.
- Built **Python-Flask** and **MongoDB** backend to store and process image and EMG data from wearable sensors.
- Utilized Mirror-JS API and Filestack API to send reminders and visual stimulus to patients on Google Glass.

Visual Question Answering

- Implemented Stacked Attention Networks in **Python** and **Pytorch** to answer descriptive questions about images using the MSCOCO dataset.
- Implemented network using a combination of **VGG16** network, **LSTM** network and attention layers.
- Model performed **10% better** than baseline CNN+RNN model.

Education

- **University of California, San Diego (UCSD),** La Jolla, CA **Expected Spring 2020**
  Master of Science in Computer Science
  Artificial Intelligence and Computer Systems

- **University of California, San Diego (UCSD),** La Jolla, CA **Fall 2014 – Spring 2018**
  Bachelor of Science in Computer Science