

SHASHWAT KAPOOR

DATA ENGINEER

CONTACT

Mobile_+1 (202) 320-9687

Email_ shashkap@umd.edu

PORTFOLIO

Github_/RaXephon

LinkedIn_/in/shashkap

Website_/raxephon.github.io

SKILLS

Languages

- Python
- C/C++
- Java/Scala
- Rust
- OCaml
- Ruby

Frameworks/Tech

- Heroku
- GC Compute Engine
- Hadoop/Spark
- Tensorflow/PyTorch
- AWS Compute
- AWS Database

BackEnd

- Flask
- PHP/jQuery
- Node.js

FrontEnd

- XML/HTML5/CSS
- Angular.js
- React.js

Databases

- SQLite/MySQL
- Microsoft SQL Server
- MongoDB

Computational Analysis

- R
- SAS

Very familiar with **C**, **Python**, **Java** and **Node.js**.

Familiar with **Bash**, **Unix**, and **Linux**.

Strong **collaboration**, **problem solving** and **critical thinking** skills.

PUBLICATIONS

- *S Kapoor, K Chang, and Y Kamariotis (2020). Image-to-Image Translation with Conditional GANs.
- S Agrawal, T Lin, S Kapoor, T Balachander, M Konduri, and B Carlisle (2018). MultiSeg: Multiple-Instance Video Object Segmentation with Unsupervised Instance Tracking.

*available upon request

EDUCATION

#University of Maryland, College Park

B.S. in **Computer Science** w/ a minor in **Statistics**

Graduation: **May 2020**

Selected Coursework: Deep Learning, Machine Learning, Design & Analysis of Algorithms, Artificial Intelligence, Cryptography, Human-Computer Interactions, Bioinformatics, Data Structures, Computer Architecture, Computer and Network Security

Extracurriculars:

- Safety Officer & Social Chair in Badminton Club, member of Culinary Club
- HopHacks 2017 (**3rd Place**), Techinca 2019 (**Winner**), MedHacks 2018 participant

WORK & RESEARCH EXPERIENCE

#Software Engineering Intern

Jun 2019 > Sep 2019

@ Fraunhofer CESE, *Riverdale, MD*

- Designed an ETL workflow for real time data processing (currently > 1 million records) using **AWS** and **PySpark** on in-house Spark buckets.
- Developed a comprehensive web based user dashboard for data visualization and predictive analysis using **Dash Plotly**.
- Implemented a **deep discriminative autoencoder** architecture for feature extraction from contracted party's **S3** image buckets.

#Data Science Intern

Aug 2018 > Oct 2019

@ Tiwary Group, *College Park, MD*

- Implemented **Monte Carlo** techniques to simulate mathematical models to study molecular interactions between protein folds and other biochemical events using **numpy** & **tensorflow**.
- Conducted computational research on modeling the interplay between **thermodynamics** and **dynamics** in complex real-world systems, relevant to **material sciences**.

#Research Assistant

Jan 2018 > Jun 2018

@ Cognitive Neuroscience of Language Lab, *College Park, MD*

- Performed data mining and data manipulation using **pandas**, **numpy** and **pyaudio**.
- Provided consistent, high quality logistic support to researchers for successful experiments.
- Assisted with setting up and conducting experiments (usually **EEG** tests).

PROJECTS (AVAILABLE ON GITHUB)

#A Wanderfull Life - Technica (**Winner**)

Nov 2019 > Nov 2019

- Built an **iOS** app with **google cloud** and **tripAdvisor**'s API as backend that allows users to safely navigate around and visit the most popular landmarks in cities via public transportation.
- Implemented **dbscan clustering** on Kaggle datasets to construct clusters of crime & identify trends of risky areas within cities and used **wmata**'s rest api to create a graph of bus stops.
- Applied **A*** with a **heuristic** function that combined distance from violence centers gathered by our model with time-cost and distance-cost to identify safest and most convenient routes through the city.

#BRAIN Project - Group Project

Aug 2019 > Sep 2019

- Worked on synthesizing images by applying transformation functions to images from publicly available datasets using **opencv** and **numpy**.
- Prepared a COCO-like dataset architecture for object detection from the synthesized images and storing them in a **aws s3** bucket.
- Developed a **convolutional neural net** for brain hemorrhage detection in CT scan slices using tensorflow and keras; training on **aws ec2**.

#Twitter Analytics - Personal Project

July 2019 > Sep 2019

- Set up a data pipeline in **java** to collate real time tweet data from the **twitter4j** library and implemented basic filtering.
- Developed a **mapReduce** job for analyzing and interpreting retweet behavior to predict trending topics in the future.