Sid Gupta

https://sidguptacode.github.io/#/
sidgupta781@gmail.com

Education

University of Toronto · Computer Science Sept. 2017 – May 2022
Bachelor's of Science · 3.5 cGPA · High Distinction

Coursework – AI: Neural Networks · Probablistic Learning · Computer Vision · Language Processing
Systems: Algorithms · Operating Systems · Computer Security · Distributed Systems · Databases

Experience

PathAI Aug 2022 - June 2023
Machine Learning Engineer PyTorch · OpenCV · Pandas · MapReduce Boston, MA

- Technical owner of **two machine learning products and two scalability projects**; products analyze histopathology images (the most common medium for diagnosing & treating cancer) and detect visual cancer features that model complex cancer environments, for diagnosis, clinical trials, drug discovery, & research
- Thoroughly explored unstructured and incomplete medical datasets. Connected metadata constraints with market use cases in-the-wild, and constructed evaluation dataset splits to ensure robustness out-of-domain
- Developed models that perform non-inferior to the average pathologist **through 80+ experiments**. Identified overfitting and complex failure modes using data visualization, pathologist feedback, & quantitative metrics. Targeted failures by sampling new input distributions, applying augmentation, tuning loss coefficients of substrata, and collecting more data in active learning
- Innovated a framework for quantifying model uncertainty using MC dropout. Engineered a MapReduce system that generates uncertainty heatmaps **20x faster than the existing solution**. Built visualizations for engineers to highlight complex failure modes & collect annotations where the model is highly uncertain

University of Toronto May 2020 - July 2022
Machine Learning Researcher Tensorflow · Keras · NumPy Toronto, ON

- **Published three papers** in machine learning, computer vision, & medical imaging with open-source code
- Built versatile machine learning systems for emulating clinical predictions on a new kind of rapid diagnostic test (RDT), applied for blood-typing and COVID-19. System first processes RDT images using OpenCV, then generates hand-crafted image features, and finally trains machine learning models using Scikit & Keras
- Developed a probabilistic algorithm for contour completion that significantly improves edge detecton quality. Integrated this algorithm with state-of-the-art image inpainting models (EdgeConnect, GatedConv) in TensorFlow, to provide additional edge context in generation, and achieved a significant SSIM improvement

The Aphrodite Project Sept 2020 - Present
Engineering Lead Docker · Scikit SDK · Pandas · React · Node Remote

- Co-engineered a web platform for students, where they can fill out a personality survey with 75 questions, and get matched with another student through a compatibility algorithm; **70,000+ users matched**
- Led the development of an automated matching codebase by defining a unified data model with Pydantic, storing & loading data compliant with that model, and running backend code in a versioned Docker image
- Trained ML classifiers to predict relationship success, using data from a follow-up survey. Applied model interpretability algorithms (SHAP, boosting) to highlight what survey answers indicate relationship success

Apple Jan 2020 - Apr 2020
Software Engineering Intern Typescript · GraphQL · PostgreSQL · React Cupertino, CA

- Built a ticketing API service that reads errors from hardware devices, and assigns tasks to internal engineers. Implemented a 'ticket search' feature with GraphQL query and mutation endpoints; enabling a client to query for tickets by attributes such as device code, unit #, build version, etc.

Skills

- **Languages**: Python · R · Julia · C · C++ · Java · TypeScript · SQL
- **SWE Tools**: Docker · Kubernetes · GraphQL · Node · Django · React · Redux · Zookeeper
- **AI / ML Tools**: PyTorch, TensorFlow, Keras, NumPy, Pandas, OpenCV, Spacy, Scikit-SDK