Srishti Sehgal

Data Scientist

613-323-4619

srishtisehgal

SrishtiSehgal

srishti.sehgal100@gmail.com

https://srishtisehgal.github.io/

Remote/Canada/USA

Contact Me

Education

University of Toronto

Engineering Science
Computer Science/Robotics Minor
Thesis: Feature Selection Analysis on PPG Data - Biometric Authentication

Skills

Languages

Python	Java	C
No/SQL	JavaScript	Bash
R	HTML/CSS	C#

Frameworks and Techologies

Torch	TensorFlow
Node.js	Caffe
React.js	Theano
Flask	Scikit-Learn
H2O	Keras
Cassandra	MySQL
Jira	jQuery
Docker	Kubernetes
AWS	GCP
Raspberry Pi	Git
Jenkins	Hadoop/Spark

Examples!

Sample Projects

*see my website for descriptions and full list of data science and machine learning topics

Publications

*see my website for a full list

Hobbies

- Dancing Bharatanatyam
- Being a Foodie
- Chasing cats

Relevant Experience

ODAIA Intelligence Inc.

Machine Learning Research Engineer / Software Engineer **May 2020 – Present**

Collaborated with the research and development team to address business problems inspired by existing literature and open-source tools. Produced ETL and ML pipelines to bring these ideas to life. Major projects include: recommendation systems, attribution modelling, causal inference, customer journey analysis, trend detection, explainable AI, end-to-end churn prediction, customer segmentation, synthetic data generation

- Created a pipeline bug queue system that sends error notifications to Google Chat from AWS to appropriate engineers, improving efficiency by 20% with quick turnaround time.
- Developed internal Python packages to reuse code for multiple projects, to ensure clean repositories
- Improved overall accuracy by 10% by installing a Git workflow protocol to streamline Git usage practices in the dev team
- Providing technical leadership, mentoring junior engineers, designed a new recruiting system for data science hires and interviewed Senior ML engineer candidates
- Driving data science efforts with in-house algorithms, developing validation procedures using literature review, open source packages and statistical theory to improve efficiency and catch bugs and engaging in back-end development/containerization to deploy these new algorithms to production
- Understanding customer requirements (time constraints, feasibility analysis) and communicating my proposals to upper management with technical presentations and demonstrations

AuToronto (Self-Driving Car)

Machine Learning Engineer **May 2019 – May 2020**

Co-led the object detection team, developed ML pipelines and implemented state-of-the-art machine learning models to analyze traffic signs and traffic lights for critical information using existing literature and open-source tools

- Conducted efficient code reviews, analyzed merge requests and presented paper summaries to all teams
- Developed computer vision models using PyTorch and TensorFlow in Python and custom-made libraries in C for faster computation. Speed improved by 20% and accuracy improved by 10% compared to previous years
- Curated unit tests for each stable release, milestones and developed internal tools/guidelines to improve code consistency, structure and interoperability
- Placed 1st in years 1&2 of the Autodrive challenge outperforming the competition by a 30% margin

National Research Council Canada

Freelance Data Scientist **September 2019 – May 2020**

Collaborated with a senior data scientist to conduct literature review, initiate source control protocols (Git) and produce an automated pipeline to bring these ideas to life efficiently and properly. 2 publications pending.

- Implemented structural dissimilarity measures, Gaussian mixture models, intrinsic dimensionality, OPTICS and PAM clustering and deep learning models to detect the effects of climate change through satellite image data with 80% accuracy
- Designed and constructed deep learning models (e.g. CNN, GAN, LSTM) with active learning to predict molecular properties and accelerate discovery of new materials

University of Toronto (Vector Institute)

Machine Learning Researcher **August 2019 – May 2020**

Investigated multiple feature selection and dimension reduction strategies, including an in-house method, to improve biometric authentication results using PPG data. Thesis work.

- Architected a private Python package to intelligently tune an ML pipeline and train a model for any time-series dataset. This allowed the research lab to automate experiments and conduct rapid prototyping 55% faster.
- Enhanced one view of data (e.g. fiducial point derivations from PPG signals) with another view (e.g. Discrete Wavelet Transform applied on PPG signals) to explore intermodal relationships and ultimately boost the model scores greatly

National Research Council Canada

Machine Learning Engineer and Data Analyst **May 2018 – September 2019**

Discussed new ideas based on literature review in a multi-faceted team and established ML pipelines to assess them. Developed a dashboard to help the client examine system health data (e.g. health of fighter jet parts subjected to full-scale testing). Published 4 papers and authored 5 technical reports.

- Achieved high accuracy with ML techniques (e.g. semi-supervised, unsupervised and supervised) and various metaheuristics (e.g. Genetic Algorithm) using TensorFlow, PyTorch and Keras; reduced training time by 85% using GPU
- Delivered high-quality presentations in weekly group meetings and conferences (e.g. CASI)
- Used test-driven development, object-oriented programming and Agile development to automate data collection and prediction from full-scale testing projects. This method provided a faster development cycle (time decrease by 80%) and easier code to maintain going forward
- Mentored junior engineer on ML projects and initiated and organized team building events, engaging ~200 people

University of Toronto

Research Assistant **May 2017 – January 2018**

Designed and established a modular suite of software and an interactive visualization dashboard using MATLAB, Python and R to perform data preprocessing, image segmentation, visualization and analyze custom properties from experimental data. Published and presented research

- Accuracy (measured by how well the software can segment the image) was between 80-95% amongst 50 experiment sets with each set containing 15 images
- Improved efficiency by 75% which also allowed the team to design more complex experiments

InkSpire

Back-End Engineer **January 2017 – May 2017**

Collaborated with 4 teammates to develop, establish a testing framework, manage and design an online text editing platform like Google Docs

- Redesigned and implemented a more user-friendly interface (A/B Testing) and accelerated response time by at least 30%
- The final project helped increase web traffic for this non-profit by 25%