## Srishti Sehgal


CONTACT INFORMATION

*Telephone:* (613) 323-4619
*E-mail:* srishti.sehgal100@gmail.com
*Website:* https://srishtisehgal.github.io
*LinkedIn:* www.linkedin.com/in/srishtisehgal


EDUCATION

**University of Toronto**, Toronto, ON, Canada

B.A.Sc Candidate, Engineering Science in Aerospace Engineering with a minor in Robotics and Business, Class of 2020

- Dissertation: Using photoplethysmogram (PPG) signals for biometric authentication with feature selection strategies in machine learning
- Advisor: Dimitrios Hatzinakos[1]

TECHNICAL SKILLS

**Languages and tools:** Python, Java, C, R, SQL, Verilog/Assembly, Windows Command Prompt, Unix/Linux shell (bash), Arduino/Raspberry Pi, MATLAB/Simulink

**Web Technologies:** HTML/CSS, Javascript, Bootstrap, React.js, Node.js

**Software Engineering:** test driven development, scalable architecture design, code review, agile development, refactorization of legacy code, source control with Git (GitLab, GitHub, AWS Code Commit), Jupyter Notebooks/IPython, vim, LaTeX, JIRA/Confluence, object-oriented programming, RESTful API, parallelization (GPU)

**Data Science/Machine Learning Topics:** linear/logistic regression, supervised/unsupervised/semi-supervised learning, deep learning (e.g. MLP, Variational AutoEncoder, CNN, RNN/LSTM/BERT), Fourier Analysis, Clustering (e.g. k-Means, h/DBSCAN, Hierarchical Clustering, OPTICS, Spectral, GaussianMixture, K-modes, K-prototypes, K-medoids, ROCK, TwoStep, Bi-clustering, subspace clustering; using quality metrics like silhouette score/dunn index/calinski index), k-Nearest Neighbours (also using FAISS, ScaNN to find look-a-likes in a dataset), Random Forest, SVD, PCA, NLP, SVM, Visualization in 2D space(e.g. t-SNE, isomap), classification/regression modelling, class-imbalance (e.g. SMOTE, over/undersampling, class-weights), multi-collinearity, imputation strategies (e.g. interpolation, iterativeImputer, fancyImputer, kNN, denoising autoencoder), validation, object detection (e.g. Self-driving car software), signal processing, parallel computing, time-series forecasting (e.g. ARIMA, exponential smoothing, Holt's Winter Algorithm), changepoint detection, trend detection, recommendation systems, explainable AI (e.g. SHAP, LIME, ELI5), time-series preprocessing (e.g. Continuous/Discrete Wavelet Transform, Spectrograms, finding fiducial and non-fiducial features, power analysis, clustering with Dynamic Time Warping), Feature selection (e.g. Multiview Single view, Dimensionality Reduction, LOF, LLE), Feature engineering, Anomaly detection, encoding techniques, CausalML (e.g. using Microsoft's DiCE package, T-/X-/S-Learners, creating control groups), Attribution modelling (e.g. using Shapley Values, first-/last-touch attribution), Target Leakage, Noisy Labels (e.g. Confident Learning), Confidence/Prediction Intervals, Developing cost functions to optimize (i.e. to replace typical ones like CrossEntropyLoss, LogLoss etc..), Churn Prediction, Simultaneous localization and mapping (SLAM - Robotics)

**Machine Learning Frameworks/Tools:** pandas, openCV, NLTK, numpy, scikit-learn, SciPy, statsmodels (library), category-encoders (library), WeKA, seaborn/matplotlib/tableau (visualization), Tensorflow, Tflearn, Torch, Caffe, Keras, Lasagne, Splunk, Hadoop

**Cloud Technologies:** AWS S3, AWS EC2, AWS Fargate, AWS RDS, AWS DynamoDB, AWS SNS, AWS SQS, AWS Lambda, AWS ECS, AWS ECR, Google Cloud Computing

**Data Storage and CI/CD tools:** pickle, HDF5, MySQL, NoSQL/Cassandra, Docker, Kubernetes, Jenkins

[1] https://www.comm.utoronto.ca/~dimitris/

## Publications

**S. Sehgal**, C. Cheung and J.J. Valdés, "Characterizing the current state of a propulsion engine: A comparison of machine learning frameworks," in *Proc. 22nd Symp. Canadian Aeronautics and Space Institute Aerospace Propulsion Symp.*, Montreal, Canada, May 10-15, 2019. Accessed on Oct. 4 2019. [Online]. Available: `https://casi.ca/resources/Documents/AERO/2019/Abstracts%20Submitted/140.pdf`

C. Cheung, **S. Sehgal** and J.J. Valdés, "A machine learning approach to load tracking and usage monitoring for legacy fleets," in *Proc. 30th Symp. Int. Committee on Aeronautical Fatigue*, Krakow, Poland, June 2-7, 2019, pp. 922-937. Accessed on Oct. 4 2019. [Online]. Available: `https://doi.org/10.1007/978-3-030-21503-3_73`

C. Cheung, J.J. Valdés, **S. Sehgal** and R. Chavez, "Failure modeling of a propulsion subsystem: Unsupervised and semi-Supervised approaches to anomaly detection," in *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol.33, no.11, Apr. 3, 2019. Accessed on Oct. 4 2019. [Online]. Available: `https://doi.org/10.1142/S0218001419400196`

C. Cheung, N. Kilfoyle, J.J. Valdés, **S. Sehgal**, R.Chavez, "Low-dimensional spaces for relating sensor signals with internal data structure in a propulsion system," in *Advances in Science, Technology and Engineering Systems Journal*, vol.3, no.6, pp.23-32, Nov. 01 2018. Accessed on Oct. 4 2019. [Online]. Available: `http://dx.doi.org/10.25046/aj030602`

R. Ranjan, **S. Sehgal**, and A. Ramachandran, "Capillarity and convection-controlled assembly during the spreading of particulate suspensions on an air-liquid interface," in *Bulletin of the American Physical Society*, 2018. [Abstract]. Accessed on Oct. 4 2019. Available: `https://ui.adsabs.harvard.edu/abs/2018APS..MARR46001R/abstract`

R. Ranjan, **S. Sehgal**, J. Kornfield, and A. Ramachandran, "Convection and capillarity induced pattern formation in the spreading of a concentrated suspension of rigid spheres over a liquid-air interface," in *AIChE Annual Meeting*, 2018. [Abstract]. Accessed on Oct. 4 2019. Available: `https://www.aiche.org/conferences/aiche-annual-meeting/2018/proceeding/paper/461c-convection-and-capillarity-induced-pattern-formation-spreading-concentrated-suspension-rigid`

## Papers in preparation

J.J Valdés, A. Tschagang and **S. Sehgal**. Developing a computationally intelligent approach to derive molecule architectures given quantum electronic properties of molecules.

J.J Valdés **S. Sehgal** and A.Pou. Designing novel machine learning architectures like generative adversarial networks (GANs) to forecast weather patterns using image processing techniques on weather satellite images.

## Conference Presentations

**S. Sehgal**, C. Cheung. 2019. A comparison between fuzzy and non-fuzzy computing methods to handle imprecision in the classification of data. National Research Council Canada Research Conference, Ottawa, Canada, August, 2019.

**S. Sehgal**. 2019. Characterizing the current state of a propulsion engine: a comparison between machine learning frameworks 22nd Symp. Canadian Aeronautics and Space Institute Aerospace Propulsion Symposium, Montreal, Canada, May, 2019.

**S. Sehgal**, S. Brunet, R. Rutledge. 2018. Inboard Leading Edge Flap (ILEF) Root Lug Calibration: Modelling the relationship between strain and load. National Research Council Canada Research Conference, Ottawa, Canada, August, 2018.

## Professional Experience

**ODAIA Intelligence Inc.**, Toronto, Canada

*Machine Learning/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

- 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 and mentoring junior engineers, designed a new recruiting system for data science hires and conducted resource interviews, resume reviews and phone screens for potential Senior ML engineer hires
- Driving data science efforts with in-house algorithms, developing validation procedures 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

**Vector Institute**, Toronto, Canada

*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**, Ottawa, 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

**aUToronto**[2] **(Autonomous Vehicle Initiative)**, University of Toronto

*Software Engineer (Object Detection System)* **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%

[2] `https://www.autodrive.utoronto.ca`

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**, Ottawa, 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
- 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
- Wrote and refactored code (Python, Java, C, C++) for a diverse array of client and internal projects

**Laboratory of Complex Fluids**, University of Toronto – Department of Chemical Engineering

*Data Analyst* **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
- Organized large amounts of information and drew conclusions from research to co-author our results on top of a full-time course load

**INKSpire**, Remote

*Back-End Developer* **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%

**SPAN Manufacturing Inc.**, Markham, Canada

*Manufacturing Engineer* **May 2016 - July 2016**

- Detailed quality inspection including dimensional analysis, Surface Mount Technology management and reading from engineering drawings of products using IPC-A- 610 Class 2 Standards to ensure integrity of company's products and end-consumer safety
- Collaborated with engineers from Autoliv Inc. and performed experiments on 220 radar sensor modules to provide detailed data and analytical research to their Global Returns Centre
- Communicated with clients to update them on the production and shipment of orders

## Relevant Projects

**Engineering Design Projects**
- Design of a remote controlled model aircraft *January 2020 - April 2020*
- Design of a lunar multiple ascent/descent vehicle for MDA Inc. *September 2019 - December 2019*
- Design and construction a fully autonomous can-recycling robot *January 2017 - April 2017*
- Design of a new skate rental storage unit at Nathan Phillips Square *January 2016 - April 2016* [3]

**Chatbot for Smart Mirror**
- Implemented several state-of-the-art speech-to-text models and optimized for high accuracy using personal audio files
- Developed a mini pipeline to query the internet with Google's API to address commands identified by model and implemented on Raspberry Pi

**Dance Critique Tool using Computer Vision**
- Implemented several state-of-the-art object detection models and optimized for 70% accuracy using personal videos of Bhartanatyam (Classical Indian Dance)
- Optimized for very low training time using GPU acceleration
- Developed ETL pipeline to use active video/streaming image data to provide the dancer a score of artistic quality

**Ad blocker chrome extension with a machine learning approach**
- Developed object-oriented code to detect spam ads on a webpage using deep learning-based models in Tensorflow with natural language processing
- Hid spam ads with 88% accuracy
- Used Github for version control and designed with test-driven development techniques

**Constructing a novel architecture for deep learning**
- Developed code for designing chemical syntheses from scratch using PyTorch and Tensorflow by starting with literature review on existing models
- Designed and ran a robust evolutionary algorithm wrapper on GPU (CUDA) around the machine learning model to optimize hyperparameters, achieve high accuracy and reduce training time by 75%

**Analysis of audio features of Spotify's hits using machine learning**
- Implemented a modular data pipeline to preprocess music features (e.g. loudness, artist) into model
- Organized several feature selection techniques to unravel importance of music features that were extracted from Spotify's REST API and maintained in a mySQL database

**Computer vision program for recyclable material detection (object segmentation)**
- Optimized a deep learning architecture in PyTorch and developed a computer vision algorithm (Python) to segment recyclable materials from non-recyclable material in order to decide which items can be recycled (e.g. metal) in a given image of household waste
- Applied ML tools to achieved 78% accuracy, under 36 hours of data pipeline development and training, and sped up the algorithm by 4X through vectorization and GPU processing

**Stock market prediction using machine learning**
- Developed a Python program in a team of two other students that predicts the movement of a user-defined stock/tweet
- Intensive pre-processing was performed on twitter data and financial open/close price data to ensure that missing terms were accounted for and that noise terms like punctuation and language stop words were removed from the dataset
- Tuned BERT model, after reading its pseudo-code from publication, in PyTorch with Tensorflow

[3] http://www.citynews.ca/2016/04/08/u-of-t-engineering-students-tackle-10-community-challenges/

backend to achieve an accuracy of 82% all in the span of 36 hours

**Car loan default prediction**

- Identified attributes having strong correlations with default target data using dimension reduction techniques (like PCA and autoencoders) and discarded irrelevant features to improve model performance by 24%
- Developed ETL pipelines using agile development with Python and mySQL and normalized all continuous data with z-scores and encoded ordinal/categorical features to improve model performance by 40%
- Explored network-based models (DNN), trees, clustering algorithms using PyTorch and scikit-learn with custom Python code to automatically tune hyperparameters. Achieved ranks within top 14% of all competitors

## Honors and Awards

Dean's Honour List
Professor Morris A. Cohen Scholarship in Engineering Science, **2018**
Eric Miglin Scholarship, **2018**
Dr. John Hamilton Parkin Scholarship, **2018**
Ryn Pudden Memorial Award, **2017**
Undergraduate Student Research Award, **2017**
Dean's Merit Award, **2015**

## Memberships

**Aggregate Intellect**, Canada
*General Member* **May 2019 - Present**

**University of Toronto Machine Intelligence Student Team**, University of Toronto
*VP External Affairs* **May 2019 - May 2020**

**Women in Stem Outreach (National Research Council)**, Ottawa, Canada
*Leading member for Logistics and Mentorship* **May 2018 - August 2019**

**University of Toronto Space Design Competition** , University of Toronto
*VP Competitions* **May 2018 - May 2019**

**Galbraith Society for Undergraduate Research Journal**, University of Toronto
*VP Finance (Awarded $1450 from grants)* **July 2017 - May 2018**

**NSight Mentorship**, University of Toronto
*VP Finance and Mentor (Awarded $1000 each year from grants)* **July 2016 - May 2018**