# Kristy Mualim

https://kmualim.github.io/

Email : kmualim@stanford.edu
Mobile : 310-904-8155


## Education

- **McGill University** — Montreal, QC
  *B.Sc in Biochemistry & Computer Science (C.S)* — *Sept. 2015 – Dec. 2018*
  - **BL21 Research Scholar**: Top 20 Independent Research Scholarship for Most Innovative Research Ideas
  - **Yale hackathon sponsor award**: Highest social good impact & most scalable use of Machine Learning
  - **Executive Roles in student groups**: Managed HANY[1], Branch-Out Mentorship[2], Roots[3]
- **International Culinary Centre** — New York, NY
  *Professional Diploma for the Culinary Arts in French Cuisine* — *Jan. 2015 – July. 2015*
  - **Maintained Perfect Attendance and Service Awards**:

## Research Experience

- **Research Assistant** — Stanford University, Palo Alto, CA
  *Principle Investigator: Dr. Anshul Kundaje, Department of Genetics & C.S* — *May. 2019 - Present*
  - **ENCODE Working Group Lead**: Leads data analysis and pipeline generation in key working group of 41 participants in ENCODE Consortium[4].
    Robust testing of validation pipeline across multiple celltypes to be used by Consortium.
    Involved in problem formulation and hypothesis testing based on different data input types.
    Frequent Presentations to non-technical audience.
  - **Data Analytics and Deep Learning**: Implemented and interpreted statistical and deep learning models on multi-model biological data inputs.
    Curated and combined datasets for exploration and downstream prediction tasks.
    Improved ease of use and validation practices of relevant code and packages.
- **Undergraduate Research Assistant** — McGill University, Montreal, QC
  *Principle Investigator: Dr. Jerome Waldispuhl, Department of Computer Science* — *May. 2018 - Jan. 2019*
  - **Structural Biochemistry and Bio-informatics Research**: Implemented CNNs to solve multiple sequence alignment problem via utilizing human-computing crowd sourcing platform - Phylo[5]
    Implemented machine learning algorithms for puzzle feature extraction based on importance.

## Projects

- **Text Classification of US Airline Twitter Sentiment Data**: Utilized data from twitter to analyze US Airlines Sentiment Analysis
  Generated jupyter notebooks for data exploration on twitter data to curate cleaner text training data.
  Implemented Text Classification Model (ULMFit) on curated twitter data to understand vital areas of improvements for future airline business models.
  Achieved 5% improvement in classifying tweets as positive/negative sentiment
- **Deep learning in gene expression inference**: Utilized deep learning to predict gene expression of target genes via landmark genes in the LINCS Consortium
  Introduced improved baseline algorithms and improved model prediction by 10%
  Analyzed and preprocessed ¿ 1TB RNA-seq, GTE and GEO expression data using Google Cloud Computing (GCP)
- **Addressing building accessibility**: Utilized machine learning on accelerometer data to measure building accessibility to advise on better building standards for all. [6]
  Integrated Google's API to generate locational heatmaps of obstacle (stairs/ramps/elevators) difficulty.
  Achieved close to 99% prediction accuracy in identifying obstacles using accelerometer data

[1] a non-profit student run organization that offers French & English tutoring to refugees.
[2] An initiative to provide after-school creative programs for students in community high schools
[3] An initiative to de-stigmatize issues on mental health. *open-mic events to be featured in Green lion films documentary*
[4] https://www.encodeproject.org/
[5] https://phylo.cs.mcgill.ca/; a crowd-computing platform for multiple sequence alignment.
[6] https://mcchillteam.wixsite.com/maxcessibility/home-1;

## Projects

- **Computer visualization & classification task**: Utilized CNNs and deep learning to predict hand-drawn images from GoogleDraw Competition, improved prediction accuracy by 11%.
  Implemented data augmentation strategies to format image data
- **PyTorch Open-source Implementation:**: Contributed cGAN implementation

## Posters/Presentations/Talks

- **Computational Validation of Enhancer-Gene Linking Approaches Using Sequence-based Models**
  *Presented at ENCODE Consortium Meeting 2019 & Stanford Department of Genetics Retreat*

## Relevant Courses

| **Core Courses** | **Other Courses** |
|---|---|
| Applied Machine Learning | Probability |
| Software Systems | Calculus & Linear Algebra |
| Computational Biology Research & applications | Statistics |
| Algorithms & Data Structures | Fundamentals of Computing |

## Programming Skills

- **Languages**: Python, Java, C, Linux BASH Scripting, R
  **Software**: PyTorch, TensorFlow, Scikit-Learn, Pandas, Numpy, Seaborn (Data Visualization), Matplotlib, Google Cloud Computing (GCP), AWS, Kubernetes, SQL, Twitter Search API