Kristy Mualim

https://kmualim.github.io/

Email : kmualim@stanford.edu
Mobile : +65 87481588

Education

- **McGill University** — Montreal, QC
  B.Sc in Biochemistry — *Sept. 2015 – Dec. 2018*
  - **BL21 Research Scholar**: Scholarship Awarded to Independent Researchers for Innovative Interdisciplinary Research Ideas
  - **Yale Hackathon Award**: Highest social good impact & most scalable use of Machine Learning
- **International Culinary Centre** — New York, NY
  Professional Diploma for the Culinary Arts in French Cuisine — *Jan. 2015 – July. 2015*

Major Research Experience

- **Research Assistant** — Stanford University, Palo Alto, CA
  Principal Investigator: Dr. Anshul Kundaje, Department of Genetics & C.S — *May. 2019 - Present*
  - ENCODE Data Analyst Lead
    - Manages key working group of 41 participants in ENCODE Consortium.
    - Built and optimized data mining and analysis pipeline, improved speedup, usability and reproducibility by 33%.
    - Spearheads robust testing of validation pipeline to be used across mutliple scientific labs in the United States.
    - Frequent Presentations to non-technical audience.
  - Predictive Analytics and Deep Learning
    - Involved in problem formulation, scoping and hypothesis testing based on different data modalities (Language, numerical data inputs)
    - Led model development, testing and validation for regression analysis.
    - Compiled and curated data inputs by web scraping using Python.
    - Implemented and interpreted statistical and deep learning models for downstream correlation and predictive analysis.
- **Undergraduate Research Assistant** — McGill University, Montreal, QC
  Principal Investigator: Dr. Jerome Waldispuhl, Department of Computer Science — *April. 2018 - Jan. 2019*
  - Human-in-the-loop (HITL) Machine Learning Research
    - Built Convolutional Neural Networks using PyTorch via utilizing a human-computing platform to solve multiple sequence alignment - Phylo; a crowd-computing platform for multiple sequence alignment
    - Performed importance feature extraction to effectively fine-tune models, improving model performance by 25%
    - Analyzed and interpreted user data to modify puzzle difficulty based on player skill
- **Undergraduate Research Assistant** — McGill University, Montreal, QC
  Principal Investigator: Dr. Kalle Gehring, Department of Biochemistry — *May. 2017 - Aug. 2018*
  - Protein Crystallography
    - Mentored undergraduate student in research project, through teaching scientific concepts and reasoning behind experimental procedure. Guided student through required experimental techniques and procedures.
    - Performed end-to-end pipeline of protein expression to purification, utilizing standard techniques (SDS-PAGE, affinity, ion exchange, size exclusion chromatography via FPLC)
    - Research documented for undegraduate research project.

Leadership Experience

- **Community Organizer** — Palo Alto, CA
  Sunrise Movement — *Aug 2019 - Present*
  - Justice, Equity and Anti-Oppression
    - Facilitated General Hub Meetings and Anti-Oppression Discussions, addressing issues like incorporating greater equitable recruitment, accessibility.
    - Spearheaded Compensation Fellowship program to diversify and introduce greater accessibility to organizing through more equitable recruitment, with a focus on recruiting Black, Indigenous, People of Color as well as working class folks.
    - Emphasized prioritization on partnering with indigenous sovereignty community organizers.
    - Working alongside with non-profits like Buy-In to bring greater awareness to existing insurance policies for communities at disaster-risk areas.
  - Voice in Local Policy
    - Partnered with Local Officials to push for more aggresive 2020 Sustainbility Goals.
    - Initiated new building regulations to transition from natural gas to electric fittings.

Key Projects

- **Improving building accessibility via Machine Learning**: Aimed to measure building accessibility to advise cities on better building standards for all. Hand-collected and curated data and built pre-processing pipelines for use with non-technical audience. Utilized accelerometer data to address obstacle difficulty using Image regression model. Integrated Google's API to generate locational heatmaps of obstacle (stairs/ramps/elevators) difficulty. Achieved close to 99% prediction accuracy in identifying obstacles.
- **Computer Vision**: Implemented data augmentation for image data formatting. Led model selection and featurization to predict hand-drawn images from GoogleDraw Competition, improved prediction accuracy by 11% from baseline models. Achieved top 10 prediction accuracy score amongst 60 teams.
- **Deep learning in gene expression inference**: Aimed to predict genome-wide gene expression using subset of key factors. Analyzed and preprocessed 1TB RNA-seq, GTE and GEO expression data using Google Cloud Computing (GCP). Curated small representative set for training, Utilized Deep CNNs to predict expression of target genes. Introduced improved baseline algorithms for better model comparison and improved model prediction by 11%.
- **Text Classification of US Airline Twitter Sentiment Data**: Conceptualized and implemented sentiment analysis tool to address consumer preference of airlines and potential areas for improvement. Generated Jupyter Notebooks for data exploration utilizing twitter data for data cleaning and processing. Implemented Text Classification Model (ULMFit) on curated data to understand vital areas of improvements for future airline business models. Achieved 5% improvement in classifying airline sentiment
- **PyTorch Open-source Implementation**:: Contributed cGAN implementation

Publications and Presentations

Mualim, K. Theunert, C. Slatkin, M. Estimation of population divergence times from SNP data and a test for treeness. *Submitted.*
Mualim, K. Nasser, J. Engreitz, J, Kundaje, A. Computational Validation of Enhancer-Gene Linking Approaches Using Sequence-based Models
Presented at ENCODE Consortium 2019

Coursework

Computer Science
Applied Machine Learning, Software Systems, Advanced Computational Biology Research
Algorithms & Data Structures I, II, Fundamentals of Computer Systems
Probability & Statistics
Probability I, II , Calculus I, II , Statistics I, II, Linear Algebra I

Technical Skills

- **Tools**: Python, Java, C, Linux, BASH, R, AWS, GCP, SQL, Kubernetes, Git
- **Packages**: PyTorch, TensorFlow, Scikit-Learn, Pandas, NumPy, SciPy, Statsmodels, Seaborn (Data Visualization), Matplotlib, Twitter API, Jupyter Notebook
- **Statistics / Machine Learning**: Statistical Analysis, Linear and Non-linear Regression, Multi-variate Regression, Clustering, Classification, Regularization
- **Languages**: English, Mandarin, Indonesian, French (Basic)