

Yule Wang, Ph. D

yulemoon@gmail.com • Canadian Citizen (TN VISA Eligible) • +1 (604) 783-2901 • www.linkedin.com/in/yule-wang-ml/

EDUCATION

- **Doctor of Philosophy, Physics,** Sep. 2015 – Feb. 2021
Simon Fraser University, Burnaby, BC, Canada
- **Master of Science, Physics,** Jan. 2013 – Aug. 2015
Simon Fraser University, Burnaby, BC, Canada
- **Bachelor of Science, Applied Physics,** Sep. 2008 – Aug. 2012
Harbin University of Science and Technology, Harbin, China

WORK EXPERIENCE

- **SoundHound** – NLP, Virtual Voice Assistants – Mercedes, Hyundai, Mobile APP and Restaurants Apr. 2022 - Nov. 2022
Machine Learning Engineer II Toronto, Canada
 1. Built and maintained a Audio-Text Command Query Noise Detector using *Distil-RoBERTa* (BERT) transformer neural network, that could filter out 40% of noisy user voice-commands and resulted in a reduction of 20% of processing time in the real-time query stream production pipelines.
 2. Defined noisy queries for production feasibility purposes and user-centric purposes. Automated a backend benchmarking analytics pipeline of production noisy queries.
 3. Developed a Rule-Based *NLP* program for recognizing valid user-commands and created a Grammar-Detector together to assist the auto-labeling of the noisy training dataset (in which only 0.5% were properly manually labeled) by implementing entity extraction, parsing and POS tagging using *spaCy*.
 4. Created and maintained a pipeline that could automatically add more rule-based patterns when new production queries comes by implementing *spaCy*.
 5. Experimented in *Bayesian semi-supervised* classification that can potentially correct the biased rule-based auto-labeling processes mentioned above.
 6. Performed text mining of the user intentions and sentence structural patterns to detect whether a user query that provides complex characteristics, such as a query that has multiple intents or a question in a query that provides multiple choices to compare between or choose from. Programmed a tool to suggest proper text regeneration and responses.
- **Applied Quantitative Methods (AQM)** – Cross-team with Best Buy Twitter Data Analytics Jan. 2017 - Nov. 2017
Data Analyst, intern Vancouver, Canada
 1. Built a content-based spam tweets filtering system using *Naive Bayesian classifier*, that reached an accuracy of 90%.
 2. Developed a sentiment analysis model to evaluate the satisfaction improvement of customers after they were being responded to by Best Buy customer service on Twitter using the *support vector machines* (SVM).
 3. Successfully classified different topics for Best Buy tweets using the *Latent Dirichlet Allocation* (LDA).
 4. Performed gender classifications of Twitter usernames and achieved an accuracy of 92%, by implementing the *semi-supervised* character n-grams algorithm.

KAGGLE COMPETITIONS

<https://www.kaggle.com/moonswords>

- **TalkingData AdTracking Fraud Detection Challenge** Mar. 2018 - May. 2018
bronze medal, top 8% - leader
Implemented the *boosting* model *LightGBM* to predict fraudulent clicks on mobile advertisements. Highly-imbalanced dataset was oversampled by using *SMOTE* and *negative-sampling* methods.
- **Toxic Comment Classification Challenge** Jan. 2018 - Mar. 2018
Implemented gated *RNN* models, *LSTM* & *GRU* to forecast multiple *NLP* toxic sentiments, such as toxic, threat or insult.

RESEARCH PROJECTS AND PUBLICATIONS

- **PhD: Statistical Modelling and Simulations of Failure Dynamics in Random Networks** Sep. 2015 – Dec. 2020
 1. Built a random graph model that studied the random failure *dynamics* in polymer *networks* in statistical physics study purposes and successfully forecast real-world polymer failure times.
 2. Established a kinetic *Monte Carlo* Python program for simulating the fracture processes that follows a Markov chain.

Wang, Y. and Eikerling, M. "Fracture dynamics of correlated percolation on ionomer networks." *Physical Review E* **101**, 042603 (2020).
(<https://journals.aps.org/pre/abstract/10.1103/PhysRevE.101.042603>)

PROGRAMMING/MACHINE LEARNING SKILLS

- **Languages:** Python, MySQL, Bash, Git. **Libraries:** PyTorch, Tensorflow, Scikit-learn, sPacy, NLTK, Pandas, flask, etc.
Skills & Tools: Kubernetes, Statistics, Machine Learning & Deep Learning, data mining, BERT, Named Entity Recognition (NER), Random Forest, Gradient Boosting (XGBoost & LightGBM), LDA, LSTM.