Mingyue (Phoebe) Li

Personal Website: GitHub, LinkedIn
412-515-7845 | mingyuel@andrew.cmu.edu

EDUCATION

Carnegie Mellon University, Pittsburgh, PA **Aug 2020 - May 2022**
Master of Science in Health Care Analytics and Information Technology
Relevant Coursework: Machine Learning With Large Datasets, Deep Learning, Web Application Development, Applied Machine Learning, Multilingual Natural Language Processing, Bioinformatics Data Integration Practicum

University of Pittsburgh, Pittsburgh, PA **June 2020**
Bachelor of Science in Mathematics-Economics
Honors: Cum Laude, McKay Prize for Outstanding Economics Undergraduate Student.

PROFESSIONAL EXPERIENCE

Amgen, Thousand Oak, CA **June 2021 – August 2021**
Data Science R&D Intern at Global Patient Safety team

- Developed Scala query to access the ARGUS database on Databricks and developed SQL query to analyze patients' data and created Proportional of Events Signal Detection Output (PoE SDO) reports across 15 months vary 8 products.
- Used different signal thresholds and detection methodologies to generate counts for three categories of adverse events. Measured the result with validated count to determine the best signal detection model with the lowest false positive rate.
- Created PoE SDO automation prototype. Led the training session for signal detection process for new hired.

Budweiser Brewing Company APAC, Shanghai, China **March 2021 – May 2021**
Data Science Intern

- Scraped the data from business review websites (DianPing.com), visualized and analyzed the customers' reviews and restaurant features by using SpaCy and Sklearn.
- Implemented recordlinkage to match the address data to the internal database and utilized Geopy to evaluate the geographic distance between two geographic coordinates.

Deloitte China, Beijing, China **Jul 2019 - Aug 2019**
Financial Advisory Intern at M&A Team

- Utilized Excel solver to analyze client's sales data.
- Designed business dashboard by using Power BI and created heatmap for business distribution on Tableau.

PROJECTS and AWARD

Utterance to Phoneme Mapping **Nov 2021**

- Predicted the phonemes contained in utterances in the test data by giving the unaligned utterances and corresponding phonemes data.
- Trained CNN+LSTM model by using nn.CTCLoss as the criterion and used beam search to decoded the predictions.
- Used Levenshtein distance to evaluate the accuracy of the prediction.

Face Classification and Face Verification **Oct 2021**

- Built ResNet18, ResNet34, and ResNet50 network by using PyTorch and trained these models to extract and represent facial features from given 4,000 persons' pictures and passed face embedding to a fully connected layer to predicate corresponding face ID.
- Evaluated similarity between given 69,037 pictures, used trained ResNet34 to extract facial features.

VirtualHug.AI Award: MIT Hacking Medicine Third Place Winner **Oct 2020**

- Designed the prototype for risk detection application for providers and caregivers. The system will collect audio data through smart home devices and analyzed patient EHR data for risk detection.

SKILLS

Programing: Python (Sklearn, PyTorch, Pandas, Numpy, Gurobi, etc.), SQL, R, STATA, Scala.
Web Application: JavaScript, Django, jQuery, S3.
Database & Big Data: SQL, Spark.
Analytics & Visualization: Advanced Excel Skills, Tableau, Power BI, Matplotlib.
Machine Learning & Deep Learning: SVM, Random Forest, K-Mean, PCA, CNN, LSTM, RNN, BERT.
Cloud Computing: AWS, Databricks.