Shan Chen

Waltham, MA, USA 02453 | +1 (507) 210-1913 | shaaan@brandeis.edu | Web: shanchen.dev | Linkedin

EDUCATION

Brandeis University, Waltham, Massachusetts **Expected: May 2022**
Master of Science, Computational Linguistics (STEM Designated, NLP + Data Science + Linguistics)

St. Olaf College, Northfield, Minnesota May 2020
Bachelor of Arts, Mathematics, Japanese, Linguistics Concentration Majors/Overall GPA: 3.7/4.0

- Honors/Awards: Cum Laude with distinctions, Japanese National Honor Society, Pi Mu Epsilon, JASSO Scholarship, National Japanese Exam Silver Prize, Henry Luce Research Grant
- Semester abroad: Nagoya University School of Engineering, Nagoya, Japan

RELATED COURSES

Algorithms & Data Structures, Applied Algorithms, Graph Theory, Neural Computation, Natural Language Annotation for Machine Learning, Object-Oriented Software Design(C++&Java), Information Retrieval, Statistical Machine Learning, Natural Language Processing, Syntactic Theory, Semantics, Phonetics

LANGUAGES AND FRAMEWORKS

Python, C++, Java, R Studio, Flutter, Firebase, Flask, HTML & CSS, NoSQL, Pytorch
Chinese (Native), English (Native), Japanese (Professional working proficiency), German (Elementary)

WORK EXPERIENCE

Data Scientist Intern, Elphi, Cambridge, Massachusetts Dec 2020 – Feb 2021

- Working on Real-time Dynamic data with Tableau and Machine learning loan prediction algorithms.
- Integrated the algorithm into the main app using GCP Pipeline to achieve systematic automation.

Software Developer, Unismart・ユニスマ, Tokyo, Japan June 2020 – Sept 2020

- Developed a cross-platform app using Flutter and Firebase framework designed to help to connect 10k+ university students during the COVID-19 pandemic.
- Support real-time push notification, messaging, coursemate matching system using Firebase API.
- Designed customer analytics functions to assess customer acquisition and app engagement gaining over 91 percent positive feedback.

Linguistics Research Assistant, St. Olaf College, Minnesota June 2018 – July 2019

- Conducted data analysis and synthesized data into a published paper (publication listed below). The Paper was published as a linguistics book chapter in 2019 by Lexington Press as: "*The Development of Willingness to Communicate in L2 Chinese Writing*".
- Generated and analyzed Japanese web sourced linguistics data for Ito, Ph.D.'s linguistics research.

PROJECTS

Washington Post Information Retrieval System Oct 2020 – Present

- Web app retrieving information from Washington post corpus with boolean and vector space search.
- Leveraged frameworks: Flask + Python front & backend implantations.

Computational Sociolinguistic Analysis of Cross-cultural Celebrity Emoji Usage March 2019 – Present

- Created machine learning models with neural networks, SVM, KNN, and random forests to predict emoji usage based on tweets' sentiment scores, implementing LSTM model summarizing the tweets.
- Analyzed celebrities' emoji use across cultures and compared the result to common users.
- Proven differences among usages, developing a classifier to predict tweet goals through emojis.

Exploration of League of Legends Professional Winning Strategies May 2020 – June 2020

- Using 10 different statistical machine learning models including, Bagging, Penalized Regression, SVM, Random Forests, and Ada Boosting, predicting win rate. Got 82% (improved by 9%) accuracy rate with the voting system among these models as results explained pro-players' scientific playing styles.