Xiaoxiang (Max) Ma

xm53@cornell.edu | (607)229-9675

xiaoxiang-ma.github.io | linkedin.com/in/xiaoxiang-ma

EDUCATION

Cornell University   **Ithaca, NY**
M.Eng. in Operations Research and Information Engineering   *Anticipated Dec 2021*
B.S. in Information Science (Data Science Concentration)   *Dec 2020*
- Major GPA: 3.93 | Cumulative GPA: 3.68
- Relevant Courses: Machine Learning for Intelligent Systems | Natural Language Processing | Data Analytics for Information Science | Data-driven Web Applications | Text Mining in History and Literature | OOP and Data Structures

SKILLS

- Programming - Python, SQL, R, SAS, Java, JavaScript
- Tools – Pytorch, Sklearn, Keras, Pandas, NumPy, Statsmodel, SciPy, Tensorflow
- Languages: English (Native), Mandarin (Native).

PROFESSIONAL EXPERIENCE

Healtius   **Remote work**
Data Science Intern   *May 2020 – Aug 2020*
- Developed reinforcement learning deep Q neural network in collaboration for a telehealth chatbot.
- Created training dataset from medical questions and patient trials in collaboration with the clinical team.
- Designed states, actions, and reward system with markov decision processes to optimize long-term reward.
- Performed sentiment analysis on user reviews from existing competitor apps.
- Web-scraped and sampled tweets related to telehealth and digital healthcare to analyze product market fit.

Marhub   **Remote work**
Data Analyst   *Jan 2020 – Aug 2020*
- Cleaned unstructured chatbot data with large proportion of missing data and language inconsistency.
- Designed data wrangling syntax for each data column collected from a chatbot question tree.
- Implemented logistic regression and random forests models to identify vulnerable refugees.
- Redesigned vulnerability screening classification procedure through statistical analysis.
- Shipped final pre-screen vulnerability model that can save screening officers' time by 25%.

iFRE   **Beijing, China**
Data Analyst Intern   *Jun 2019 – Aug 2019*
- Implemented portfolio risk management and credit scorecard model using Python and SAS.
- Cleaned and combined 26 datasets spanning 12 million user data using SQL and SAS.
- Performed feature selection using statistical analysis including WOE, IV, Chi-square, and entropy.
- Designed logistic regression model to identify defaulted borrowers and achieved accuracy of 80%.
- Validated an existing Alipay scorecard model with company data and presented to management.

INDEPENDENT RESEARCH

Predicting Startup Longevity   **Ithaca, NY**
Independent project   *Feb 2020 – May 2020*
- Identified key indicators of longevity for startups using logistic model and data collected from angelist.co.
- Published an article featured as top story on TDS home page for 2 days with 9k+ views cumulated.

TEACHING ASSISTANTSHIP

Cornell CS4780: Machine Learning for Intelligent Systems   *Aug 2020 – Dec 2020*
Cornell INFO2950: Introduction to Data Science   *Jan 2020 – May 2020*
- Held office hours and supported students weekly on course materials, homework, and projects.