Shuhan Zhang

608-358-8668 | shuhanzh@andrew.cmu.edu | Pittsburgh, PA

https://www.linkedin.com/in/shuhan-zhang-752918179/

EDUCATION

Carnegie Mellon University, Heinz College — **Pittsburgh, PA**

M.S., Information Systems Management - Business Intelligence and Data Analytics — Dec 2022

Coursework: Big Data, NLP, Machine Learning, Data Science for Product Management, Database Management — GPA: 4.0/4.0

University of Wisconsin-Madison, College of Letters & Science — **Madison, WI**

B.S., Statistics & Mathematics — May 2021

Honors: Dean's List Year 2018 and Year 2020; Distinction Scholarship — GPA: 3.88/4.0

WORK EXPERIENCE

Accenture (China) — **Shanghai, China**

Strategy & Consulting Intern — June 2020 - Aug 2020

- Executed data **merging**, data **cleaning**, **feature engineering** to generate an integrated and clean data preprocessing result which saved **2x** time for predictive sales model building
- Conducted **SWOT analysis** and **Comparison analysis** for market competitiveness of an **IoT** client company
- Facilitated with the marketing strategy formulation and evaluation and delivered the result

Zhongding Information Technology Company — **Anhui, China**

Marketing Analyst Intern — Dec 2019 - Jan 2020

- Applied **A/B Testing** and **Regression** to provide quantitative business insights on product pricing and marketing strategy
- Collectively built a **relational database** using **SQL** to organize sales data for the convenience of future reference
- Improved **time efficiency** by speeding up 3 times the completion time of downstream task

PROJECT EXPERIENCE

Social Media Analysis of Samsung Galaxy S8 vs. iPhone 8, iPhone X — Spring 2022

- Cleaned text data by removing stop words and irrelevant tokens like hashtag sign, email, URL through **nltk** and **spacy**
- Performed **sentiment analysis** using **Vader** to compare sentiment distribution regarding Samsung versus iPhone
- Determined top attributes for both negative and positive sentiment for each product using **POS Tagging** and **Counter** to offer product pre and after launch adoption findings and recommendations for better product management

Machine Learning Based Policy Refinement to Improve Health Outcomes — Spring 2022

- Employed **k-means clustering** to group different counties with the similar health conditions and behaviors
- Tuned and compared **multi-linear regression** model and **random forest regressor** for higher prediction accuracy and **feature importance** selection
- Furnished **policy making insights** for specific counties to pinpoint key areas which could result in better health outcomes such as reducing premature death cases and improve general mental health levels etc.

Development of Classification Algorithm for Cardiovascular Disease Detection — Fall 2020

- Implemented Exploratory Data Analysis (**EDA**) to unveil underlying data pattern and potential feature relationships
- Led a team of 5 to develop a classification model with **random forest classifier** to predict the probability of getting cardiovascular disease through **model selection** by **cross validation** and **grid search**
- Improved the accuracy of cardiovascular disease detection **by 10%** from the most recent result

Yelp Rating Predictions Project and Competition Using NLP and Regression — Spring 2020

- Utilized **TF-IDF scores** and **sentiment analysis** to tokenize words and select the most valuable ones as features
- Established a multi-linear regression model by combining **PCA** to predict Yelp ratings based on the tokenized words
- Achieved **top 5** among 25 groups in accuracy competition on Kaggle for the prediction of a total 50,000 restaurants

SKILLS

- Programming: Python, SQL (Oracle, MySQL, PostgreSQL), R, Hadoop, Spark, PyTorch, Tableau, Power BI, Stata, Excel
- Data Science: Machine Learning, Regression, Big Data, A/B Testing, NLP, Deep Learning, Hypothesis Testing
- Language: English, Mandarin, Korean