bni@nd.edu

BO NI

(574)340-6550

Data Scientist

227 Siegfried Hall, University of Notre Dame, Notre Dame, IN 46556

EDUCATION

University of Notre Dame *August 2017 - May 2021*
Bachelor of Science in Mathematics and Computing Overall GPA: 3.98 / 4.0
Second Major: Economics Dean's List All Semester

COURSEWORKS/SKILLS

Programming: Fundamentals of Computing, Data Structures, Algorithms, Mobile Development, Web Development
Statistics: Time Series Analysis, Econometric, Probability Theory, Statistical Inference, Survival Analysis
Machine Learning: Support Vector Machine, Random Forest, Computer Vision, Anomaly Detection, Social Sensing

HONORS

Best External Source Prize ASA DataFest 2019, University of Notre Dame
Whitman Family Fellowship Summer 2019, University of Notre Dame

WORK EXPERIENCE

Lab of Medical Image Computation, Massachusetts General Hospital *May 2019 - July 2019*
Data Science Researcher *Boston, MA*

- **Computer Vision:** Conducted research on feature extraction with the state-of-art model VQ-VAE on NIH chest x-ray dataset which includes 11,000 images. Held seminars and presented updates during weekly meeting
- **Anomaly Detection:** Proposed an anomaly detection method using mean-squared-error obtained from a restricted Autoencoder. Model reached 82% recall rate and 0.85 AUROC score in the task of Pneumothorax detection.
- **Frameworks and Models used: Pytorch, Keras, Autoencoder, DenseNet**

Mobile Computing Lab, University of Notre Dame *September 2018 - May 2019*
Purple Martin Project Team Leader *Notre Dame, IN*

- **Software Engineering:** Developed a mobile app with Ionic and React for Purple Martin Conservation Association that will cover 3k to 4k participants a year. Functions including reports submission and news updates.
- **Team-work:** Held weekly meetings with Prof.Pollabeur and team members. Discussed needs and updates on a monthly-base with the leader of the organization (Purple Martin Conservation Association).
- **Frameworks used: Ionic, Parse, React JS**

Medical Big Data Department, Tencent *June 2018 - July 2019*
Data Analyst *Shenzhen, China*

- **Natural Language Processing:** Extracted smoking habits from over 10000 patients' self reports using snowNLP, a Chinese word segmentation tool. Analyzed the data with Latent Dirichlet Allocation to extract semantic topics.
- **Machine Learning:** Utilized bagging methods to achieve a 95% accuracy in the task of lung disease prediction.
- **Models Used: Random Forest (xgboost), SVM, snowNLP**

PUBLICATIONS

Enhancing Early-Stage Fraud Detection by Behavior Forecast, KDD20 *September 2019 - Present*
On-going *Notre Dame, IN*

- **Dynamic Origins:** Reconstructed the idea presented in the paper *Dynamic Origins of Distribution Functions* by Python to model the social network changes in Tencent Weibo from 2012-2015.
- **Big Data:** Worked with more than 500G tweet-like data. Labeled the data with a supervised algorithm with mentor. Reached 90% accuracy in a sample of size 205 that has been hand-labeled.
- **Language/Packages used: Python, Scikit-Learn**

iPoemRec, ASONAM 2019 *September 2018 - May 2019*
Second Author *Notre Dame, IN*

- **Web Scraping:** Scraped Chinese poems from online using Scrapy. Built multiple baseline systems with Word2vec and SentiBank which reached 0.79 for crowd sourcing precision.
- **Publication:** Second author of paper: *Through The Eyes of A Poet: Classical Poetry Recommendation with Visual Input on Social Media* (Accepted by ASONAM 2019 [Acceptance Rate: 12%]).
- **Language and Framework used: Python, Flask, Sentibank, Word2Vec**

ADDITIONAL INFORMATION

Languages Chinese(Native), English(Professional), Japanese(Conversational)
Interests Investment, Basketball, Soccer, Philosophy