Xuan Wang

Austin, TX78731 | Tel : 512-200-6763 | Email : xxxuan0213@utexas.edu | LinkedIn | GitHub | Portfolio

SUMMARY

Research Experience: About 2-year undergrad & graduate research experience building models and solving analytical problems using ML, DL such data-science related methods;

Skills & Tools: **Programming Languages**: Python (libraries: NumPy, scipy, pandas, matplotlib; packages: sklearn, Keras, TensorFlow), Java|**Databases**: SQL |**Visualization**: MATLAB, Tableau |**Statistics**: SPSS, R |**Frameworks**: PHP, CSS

EDUCATION

MS In Information Studies - The University of Texas at Austin | GPA: 3.85/4.00 Aug 2019 - May 2021

Courses: Data Mining, AI in Health, Database Management, Data Wrangling, Data storytelling,

BS In Electronic Commerce - Dalian University of Technology | **GPA**: 3.30/4.00 Sep 2015 - Jun 2019

Courses: .NET Programming, Java and Object-oriented Programming, Data Structure, Probability and Statistics

National Taiwan University of Science and Technology, 2016 Fall semester Exchange program | **GPA**: 3.60/4.00

INTERNSHIP EXPERIENCE

The University of Texas at Austin - Red McCombs School of Business Aug 2020 - Jan 2021

Teaching Assistant of BIG DATA & DISTRIBUTED PROGRAMMING

- Executed operations and actions of RDD in **Apache Spark** framework, to transform and merge distributed dataset.
- Implemented traditional **Machine Leaning** methods to perform statistical analysis and prediction. Built neural network models using TensorFlow to solve QA problems.
- Achieved high-performance computing and accelerate Machine/Deep learning approaches with **Amazon EC2**.
- Applied problem-solving skills to troubleshoot code issues deeply. Coordinated between different roles.

Dayi Technology Co., LTD (China) - *Database Maintenance Assistant* Jul 2018 - Aug 2018

- Created SQL servers on virtual machines to consolidate SQL infrastructure and reduce memory usage.
- Utilized primary/secondary synchronization to coordinate master and slave MySQL databases.
- Updated databases on the secondary database (VM), and fetched the modified database on the primary side.

PROJECTS

Clinical Narrative in Apache cTAKES (NLP project, focuses on clinical care) Apr 2020 - Now

- **Data Extracting & Aggregating**: Used Apache cTAKES to extract information from 30, 000+ electronic medical records. Built pipeline to fetch required core semantic concepts by extracting and transforming the obtained result.
- **Data Analysis**: Applied NLP algorithms to detect patterns in clinical notes to provide data-driven clinical decision making.

Explore the Deep learning Models with Extrasensory Dataset Feb 2020 - May 2020

- **Feature selection**: Used Sequential Forward Selection (SFS) and Auto-encoder to select features from the datasets.
- **MLP Model:** Developed traditional supervised learning methods like the Random Forest, and Neural Networks like MLP, RNN, and LSTM models.
- **Optimization**: Applied multiple tuning methods like dropout, batch normalization, to improve the balanced accuracy up to 89%.

Readmission Prediction for Hospital Feb 2020 - May 2020

- **Data preparation**: Balanced data with under-sampling and evaluated the probability distribution of words with Zipf's law and prepared cleaned labels.
- **NLP**: Represented text features with the Bag-of-Words approach, split text into chunks and created vectorizers on the clinical notes as the input features of the predictive models.
- **Modeling**: Forecasted the boolean results with regards to the input features via Random forest, CNN with LSTM and XGBoost.

Web pages Design Oct 2019 - Dec 2019

- **Framework**: Collaborated with the team to draft the layout and created table structures in terms of the entity relationships.
- **Front-end development**: Generated dynamic HTML pages (PHP & CSS). Output results in tabular forms and created pagination.
- **Database**: Added MySQL connections via PHP to query among massive tables in MariaDB.

An Intelligent Traffic Light System Based on Digital Infochemicals (Eclipse SUMO) Mar 2010 - Jun 2019

- Constructed an intelligent traffic simulation model from the perspective of complex adaptive.
- Built and conducted the simulation intersection through Traci API in python, and finally improve vehicle flow efficiency by 30%.