Xuan Wang

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

SUMMARY

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

Programming: Python (Scikit-learn, NumPy, Pandas), MySQL, R (tidyverse, ggplot2), Spark, SPSS, Matlab, PHP, CSS

Tools: Tableau, AWS, GCP, Google Analytics, Excel

EDUCATION

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

Courses: Data Mining, Machine Learning, 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

Exchange program - National Taiwan University of Science and Technology | GPA: 3.60/4.00 Sep 2016 - Jan 2017

PROFESSIONAL EXPERIENCE

Texas Department of Information Resources - Research Intern Dec 2020 - Now

• Connected with Salesforce databases by using Python REST API client to obtain the latest contract data.

• Developed a multi-threaded automated program to scrape required files from 7 different websites and handle exceptions.

• Leveraged requests, BeautifulSoup, and selenium for request sending, web page parsing and data access.

• Hold training sessions to illustrate steps to field supervisors and improve this program based on feedback.

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

Teaching Assistant of *Big Data & Distributed Programming*

• Assisted and executed operations of RDD in Apache Spark framework to transform and merge distributed datasets.

• Achieved and accelerated high-level APIs like TensorFlow approaches via AWS EC2.

• Performed high-performance deep learning APIs like Keras to enable fast experimentation.

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

• Created primary/secondary replication architectures based on bin logs within Docker and increased CPU utilization by 36%.

• Utilized Read/Write Splitting to write in the primary databases and query in the secondary databases via Mysql Proxy.

PROJECTS

Clinical Narrative in Apache cTAKES (NLP project) Apr 2020 - Sep 2020

• Used Apache cTAKES to extract, transform, and aggregate 30, 000+ electronic medical records; fetched required core semantic concepts by applying Named Entity Recognition.

• Preprocessed the unstructured data using the Bag-of-Words approach and fitted a LSTM model (reaching the accuracy of 82%) to detect possible diseases and provide data-driven clinical decision-making.

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

• Collected and transformed 225 raw features from 60 users and balanced the multi-labeled content by using Sequential Forward Selection and Auto-encoder methods.

• Constructed Random Forest, Multilayer perceptron, and RNN to compare the differences between the various models.

• Tuned weighting parameters and applied dropout and batch normalization to improve the balanced accuracy up to 89%.

Readmission Prediction for Hospitals with MIMIC-III Feb 2020 - May 2020

• Balanced data with undersampling algorithms and evaluated the probability distribution of words with Zipf's law.

• Represented chunked key features using NLTK corpus and StanfordNLP python library; Applied grid search and stratified k-fold cross validation to find the optimal hyperparameters.

• Forecasted boolean targeted labels by constructing Random forest, CNN with LSTM, and XGBoost models.

The University of Texas at Austin, Department of Computer Science Jan 2021 - May 2021

Teaching Assistant of *Cloud Computing*

• Programmed AWS S3 and DynamoDB handlers to manipulate distributed data with AWS SDK for Python (Boto3).

• Deployed Helm charts on Google Kubernetes Engine by using a single node GKE cluster to explore container systems.

• Set up a CI/CD environment using Jenkins and Bitbucket for a stateless containerized application.

PHP Web-based application Design with MariaDB 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.