Yanxiang Zhou

Atlanta, GA • (404) 218-5098 • yzhou791@gatech.edu

EDUCATION & HONORS

Georgia Institute of Technology Atlanta, US

- ***Master of Computational Science & Engineering,*** Cumulative GPA: 4.0 / 4.0 08/2021 – 05/2023
- **Related Coursework**: *Graduate Algorithms; Data and Visual Analytics; Data Mining; Parallel Computing; Operating System*

Central University of Finance and Economics Beijing, China

- ***Bachelor of Economics,*** Overall GPA: 3.8 / 4.0 09/2017 – 06/2021

SKILLS

Language:

- **Software Development:** Python, C++, Shell
- **Data Engineering:** R, SQL, Scala, MATLAB
- **Web Client Development:** JavaScript (d3.js, node.js)

Unix/Linux Environments: Ubuntu Bionic 20.04 LTS

Frameworks and tools: Pytorch, TensorFlow, Spark, Hadoop, AWS (S3, ECR), Google Cloud Platform, Tableau, Git, Spacy, Huggingface, Transformer, Flask, GraphQL, MySQL, SQLite, PostgreSQL, HTML5, CSS, JSON, OpenMPI, Bash, Docker, Jenkins, MongoDB, React.js

Analytical: Data Analysis, Data Visualization, Machine Learning, NLP, Optimization, Time Series Analysis, Operations Research, Hypothesis Testing

PROFESSIONAL EXPERIENCE

Mentra 10/2021 – 03/2022

Data Scientist Atlanta

- Working in data science team to construct a job matching system for neurodiverse
- Developed and fine-tuned **BERT** model in **Pytorch** to classify texts in job postings and created a process to automatically construct training dataset
- Trained robust **BiLSTM-CRF** model to extract job titles, locations**,** skills and proficiency requirement information from job descriptions
- Built an API on top of the job information extraction model and **GraphQL** backend, implemented a GraphQL server controller with various queries, types, mutations, and resolvers, and created a **React** front end

Peakview Capital 10/2020 – 01/2021

Data Scientist Intern Beijing

- Designed and managed an E-commerce company database using **MySQL**, queried MySQL database queries from python using Python-MySQL connector and MySQL dB package to retrieve information and analyze product types, prices, and sales
- Utilized the Gaussian Mixture Model-Hidden Markov Model (**GMM-HMM**) to predict future revenues and sales trends in target corporation
- Trained and fine-Tuned the **BART** Large Model for text summarization of investment reports and achieved Rouge-1 score 42.3, applied topic model (**LDA**) on the results to classify reports
- Deployed the NLP model on Python based **API (RESTful Web Service)** using **Flask**, **HTML**, **CSS**, **JSON**

Huaxia Bank 01/2020 – 06/2020

Software Engineer Intern Beijing

- Data processing, data analysis and data cleaning for default loans utilizing Python (**PySpark**, **NumPy**, **Pandas**), visualized data using **Matplotlib** and **Seaborn**
- Participated in building and launching a loan fraud detection system based on **xgboost** and Generative adversarial network (**GAN**) anomaly detection algorithms, used feature engineering to improve detection rate by 4.2%
- Implemented a Continuous Integration and Continuous Delivery **(CI/CD)** pipeline for the loan fraud detection system with **Docker**, **Jenkins** and **GitLab**
- Trained and fine-tuned models including **BERT** and **Text CNN** (in **TensorFlow**) on social media comments regarding banking products and services to perform the task of financial text sentiment analysis and key word extraction
- Built RESTful API for the NLP model, executed REST API call on **Apache Spark,** and deployed RESTful API to **Google Cloud Platform**

RELEVANT PROJECTS

Recipe Recommendation Website, -*Full Stack* *Web Service Development* 08/2021 – 12/2021

- Built a recipe recommendation website which allows users to choose partial ingredients and receive recommended recipes
- Developed responsive web pages using **HTML** and **CSS,** leveraged tools including **d3.js** to visualize recommendation output and create interactive charts and graphs on website
- Constructed a database of 230k recipes, performed feature engineering using **TF-IDF** and lemmatization, trained a **xgboost** regression model to simulate the relationship between ingredient combination and potential recipe rating
- Integrated the machine learning model to the web application using Flask

Text Classification and Fraud Detection for Risk Management, *China Computer Federation, Big Data Contest (Ranking: **2**/1998)* 11/2020 – 12/2020

- Applied **xgboost**, **GBDT**, **catboost** and **Random Forest** to predict loan default, performed model evaluation and selection based on AUC score, the final model achieved AUC score 0.89

  Devised an unsupervised data pre-process method (**PCA**, **K-means**) to detect missing labels, trained multi-label text classification model based on **RoBERTa**, achieved accuracy 0.93