

## OBJECTIVE: Full-time Data Scientist

### EDUCATION

#### Rochester Institute of Technology

Rochester, NY, USA

M.S. in Computer Science

8/2018 – 5/2021

*Advanced Graduate Certificate in Big Data Analytics*

*Relevant courses: Data Structures and Algorithms, Object-Oriented Programming, Intro to Big Data, Data Mining,*

*Database System Implementation, Web Service & Service Oriented Computing, NLP Text Mining*

M.S. in Game Design & Development

8/2016 – 8/2018

#### Chengdu University of Information Technology

Chengdu, SC, CHINA

B.Eng. in Electronic and Information Engineering

9/2012 – 5/2016

### RELEVANT EXPERIENCES

#### Alibaba Taobao Ecommerce User Behavior Analysis

- Conducted a quantitative data analysis for datasets including developing retrieval, processing, analysis, and visualization
- Designed and implemented the architecture data annotation pipeline with Python, Apache Spark including Spark SQL and MLlib
- Utilized the Ecommerce statistical and rule-based metrics RFM model, AARRR framework including DAU, CVR, retention rate and churn rate to evaluate the business operating using variant visualization methods with matplotlib, seaborn and plotly
- Created the annotation for user journey purchase behaviors using aggregation and KMeans Machine Learning clustering method to develop the segmentation of purchase behavior to enhance the data-driven marketing decision

#### Text Mining towards Yelp Business Category Prediction

- Built the Natural Language Processing predictive models and developed Machine Learning and Deep Learning pipeline
- Conducted data preprocessing on business and user tags by indexing, joining, groupby, aggregation with, Pandas, Numpy
- Cleaned the customer reviews for stopwords removal, text lemmatization and stemming using NLTK Stem, NLTK Corpus Reader
- Extracted textual features from customer reviews using TF-IDF, LDA, Word2Vec, GloVe with Scikit-learn, Gensim, Stanford Glove
- Implemented Machine Learning models like Logistic Regression, Support-vector machine, Naïve Bayes, Random Forest, Deep Learning models like Convolutional Neural Networks, GRU & LSTM encoder-decoder with Scikit-learn, TensorFlow Keras

#### Teen Gaming and Civic Engagement Survey Data Analysis

- Implemented a quantitative research based on large-scale questionnaire data collected by Pew Research Center
- Created the groupings and label scheme for the teen respondents by implementing Machine Learning Agglomerative Clustering method with python scikit-learn and Scipy frameworks based on the subtable of 12 game genre attributes
- Visualized the analyzing process with dendrogram, heatmap and histogram with Matplotlib and seaborn libraries

#### Online Real time YOLO Object Detection Web Service (<http://3.16.135.19/home>)

- Implemented an end-to-end object detection API to detect traffic signs and stop signs from street images with Python TensorFlow
- Built the predictive Deep Learning model based on the state-of-art YOLOv4 algorithm with Google Open Image dataset
- Developed the web page with Flask to demonstrate detecting results with scores and bounding boxes
- Deployed the web service on AWS EC2 with Unicorn, Nginx for load balancer, WinSCP, PuTTY for instance SFTP/SSH connection

### CERTIFICATES

**Big Data Analytics Advanced Graduate Certificate** (issued by Rochester Institute of Technology)

**TensorFlow Developer Certificate** (issued by Google TensorFlow Certificate Program)

**DeepLearning AI. Deep Learning Specialization** (issued by Coursera)

### TECHNICAL SKILLS

**Language:** Python, Java, C/C++, C#, Clojure, HTML, CSS, JavaScript, XML, SQL

**Database:** MySQL, MongoDB, Apache CouchDB

**Framework:** Spark SQL, MLlib, Scikit-learn, Pandas, NumPy, SciPy, TensorFlow, Keras, Matplotlib, NLTK, Genism

**Tool & Concept:** AWS, Git, SourceTree, Bitbucket, GitHub, REST, SOAP, MVC, OOP, API, CI/CD, Agile, Cloud Computing