Zhi Zhang
Data Scientist

Boston, MA - Email me on Indeed: indeed.com/r/Zhi-Zhang/fe1f9eda5b6fff93

Young Professional with Computer Science Master and five years' big data experience, looking for mid-level Data Scientist position. Number- and data- driven with strong statistical, machine learning and enterprise data application deployment skills. Dedicated team worker with effective oral, written, and communication skills. Good organizational skills- detailed oriented; passion for accuracy & follow-through. Self-motivated and driven; high personal standards & work ethics - persistent, assertive.
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist
Brigham and Women's Hospital - Boston, MA - June 2015 to Present

Top-ranked hospital nationally, and Harvard Medical School's second largest teaching affiliate

• Deployed the healthcare big data platform including architecture and technology selection using common open-source framework Spark and Hadoop, helped maintain the scalability, robustness, effectiveness, security of the platform to make a scalable storage and parallel processing of data
• Conducted data ETL, interacted with source system to understand the data and layout the specification to bring the data into the landing area. Enforced best practices and guiding principles around data acquisition and data quality.
• Linked, sorted, annotated, augmented of massive disparate, potentially incomplete and noisy big genomic, electronic medical records, Nex-gen sequencing, clinical procedure and claims data into tables for queries
• Developed and implemented NLP algorithms to perform segmentation, lexical analysis, natural language parsing, and conceptual indexing to extract accurate medical concepts from narrative notes to identify and classify patients with specific diseases and status with a high positive predictive value
• Investigated and evaluated methodologies and algorithms for solving high-dimensional data prediction, such as dimensionality reduction, tensor factorization, and regularized regression models: elastic net, sparse modeling (e.g. Lasso), Ridge, and application in large-scale patients' phenotypes predication
• Constructed different predictive models on Spark to improve computational efficiency and reduce false-positive edges, and applied it in large-scale inference of disease risks on individual patients
• Identified different medicine adherence threshold that differentiates patients regarding the clinical resources and health outcomes by using survival trees, random survival forests with application in bootstrap
• Applied recurrent neural network with gated recurrent units to predict the clinic events of patients, evaluated different hidden layers and evaluated the performance by precision, recall and F1 value

Data Science Consultant
FocusKPI - Needham, MA - June 2014 to May 2015

A leading big data analytics firm that enables its clients to make confident data-driven decisions.

• Worked on multiple big data SaaS projects related to information retrieve, pattern recognition based on internet structured and unstructured data, performed real-time algorithm, dynamic prediction, and data visualization
• Implemented computational advertising and personalized recommendation system, performed content based and collaborative filtering for item recommendation, built vector space model and performed latent semantic

analysis to match the advertisement with webpage for accurate online advertising, and predicted the CTR for internet-based retail company
• Scraped Harvard Business Review online contents, NLP of contents to quantify words, topics, applied logistic regression, GTB, Bayes, and recurrent neural network to predict online subscription rate by features such as article page view counts and article major attributes such as topic, word count, images, bullets, insets, etc., evaluated the performance of different algorithms through cross-validation
• Used Web Spider technology to pull the information together from hundreds of career web pages. Data and information retrieval by using pattern recognition, clustering, and feature selection to explore the pattern and predict whether the candidate fits and employee leave company events to support job seekers-company match
• Implemented, and maintained analytical team's enterprise data platform. Automated and optimized architecture to run heterogeneous models using Python, R, Shell Scripts and Crontab on Linux server, wrote back-end Python code to generate JSON files for data transmission, deployed the application running on virtual Linux servers

Predicative Analyst

Teletech Insights - Burlington, MA - February 2013 to May 2014

A Business Analytics company which provides data-based analytic services to Fortune 1000 companies.

• Developed multiple predictive models to solve business questions related to marketing and finance
• Fitted ARIMAX model using marketing channel spend and revenue data to forecast channel level revenue for different prediction duration & skip months' scenarios, evaluated multiple prediction methods such as propensity score, ARIMA time series, robust regression, optimization model
• Built a classification tree based segmentation framework to find out which factors most strongly affect customer's response to a promotion to identify best customers, marginal buyers and non-buyers for targeted marketing
• Built an ADL model and calculated the Adstock effect to make recommendations on advertisement spend allocation
• Built different models such as Markov chain, logistic regression to predict customer life time value

SCHOOL PROJECTS

Human-like knowledge based AI development: Designed and developed a human-level, human-like intelligent agent that can answer human intelligence tests such as visual analogy problems based on both verbal and visual representations. (Git source code: https://github.gatech.edu/zzhang608/KBAI/blob/master/CS763716/Project%203/Agent.py)

Simulated Intelligent Robot Tracking Agent development: Developed an intelligent agent that can track a simulated robot motion by sensing of the robot moving measurements through localization, Kalman Filters, searching, and control (Git source code:
https://github.gatech.edu/zzhang608/CS8803-AI4R/blob/master/projects/runaway_robot/studentMain4.py)

Residual Neural Networks for Trajectory Prediction: Developed an algorithm to predict the future trajectory of a Nano robot, evaluated multiple algorithms, reduced video data dimensionality by PCA, tuned residual Neural Network Hyperparameters and applied Bootstrap Aggregation with Multiple Residual Neural Networks

Developed and deployed the decision support learners by evaluating different Machine Learning algorithms, Linear regression, KNN regression, Bagging, Q-learning, and ensemble learning to predict the stock price and decide real time long/short strategy and allocation of investments resource, to quantitatively maximize portfolio value, provide recommendations to hedge fund and wealth management

EDUCATION

Master of Computer Science in Computer Science
Georgia Institute of Technology
September 2017

Master of Science
University of Kentucky
December 2012

Bachelor of Engineering in Engineering
China Agricultural University
June 2009