YUHAO ZHOU

●517 W 113th ST, New York, NY, 10025 217-974-0469 yz2844@columbia.edu

EDUCATION

Columbia University New York, NY

M.S. Data Science Dec 2016

Courses: Algorithms for Data Science, Statistical Inference and Modeling, Machine Learning, Computer System for Data Science, Data Visualization, Translational Bioinformatics, GIS and Spatial Analysis, NLP

University of Illinois at Urbana-Champaign Champaign, IL

B.S. Electrical Engineering May 2015

Courses: Big Data, Data Mining, Database, Data Structure, Linear Algebra, Text Information Systems, Numerical Analysis, Computer Engineering, Probability with engineering applications, Robotics, Control Systems

SKILLS

Python, Jupyter, Numpy, Scikit-learn, R, Spark, AWS, C++, SQL, C

WORK EXPERIENCE

Audible, an Amazon Company Newark, NJ

Data Scientist Intern May 2016 – Dec 2016

- Developed a model for determining the financially successful books based on intrinsic book quality using gradient boosting
- Worked on customer segmentation based on customer behaviors using hierarchical kmeans
- Text mining and NLP analysis on over 100K titles
- Used multiple data management tools including AWS S3, ETL Manager, Redshift, SQLDeveloper
- Exploring the functionality of Spark and discovering the use cases of Spark for the team

State Farm Insurance Champaign, IL

Data Scientist Intern May 2015-August 2015

- Developed an in-car Gesture Recognition System for classifying distractive driving behaviors in a multidisciplinary team
- Collected 80000 frames of 3D data of distractive behaviors of 82 drivers using a specially designed prototype device
- Implemented a KD-Tree based interpolation method to align and refine data points for the use of Random Forest classifier
- Achieved an average accuracy of 81 percent for classifying multiple gestures

DATA SCIENCE RESEARCH AND PROJECTS

Kaggle in class Predictive Modeling Competition (team of three people) April 2016 – May 2016

- Developed a binary classifier using random forest algorithm which accurately predicted the coreference of conversations
- Used one-hot encoding to process categorical variables and grid search to tune the model parameters
- Achieved an accuracy of 0.95 in the private data points

Entity Resolution Competition (team of three people) March 2016-April 2016

- Developed a classifier using random forest which determines whether the given pairs of venues matching or not
- Designed a method shrinking data dynamically to decrease running time largely by avoiding pairwise comparison of all venues
- Achieved test precision of 99.16% and recall of 97.92%

International Research Experience Program Darmstadt, Germany

Undergraduate Researcher at Knowledge Mining & Assessment Group June 2014-August 2014

- Designed a model that simulates the grading system for evaluating soccer player performance of a credible sports magazine and predicts a grade for the next match
- Built a model based on Bayes theorem using Python and used performance data of all players of Bundesliga for 2011 season

Data Science Course Projects (Python, SQL) Urbana, IL

- Built a food recommendation system with collaborative filtering for a food ordering Website Feb 2014-May 2014
- Used the GPS data to set up statistical model for finding the epicenter and occurrence time of earthquake
- Used SIFT for detecting distinctive image features

HONORS

- TU Darmstadt Scholarship of Excellence, May 2014
- Special Recognition for Outstanding Service, Leadership and Commitment to IEEE-USA and the Profession, Spring 2013
- Phi Eta Sigma National Honors Society member, Eta Kappa Nu (HKN) member
- Dean's List: Fall 2011, Spring 2012, Spring 2013, Spring 2014; James Scholar 2011, 2012

ACTIVITIES

Institute of Electrical and Electronics Engineers, (IEEE) Conference Committee Urbana, Sep 2012-May 2013

- Organized for the Student Professional Awareness Conference 2013 and IEEE Region 4 Leadership Conference 2012