DONG ZHANG, Ph.D.

US permanent resident | Seattle, WA
(614) 397-5173 | dongzhanghz@gmail.com
github.com/dongzhang84
linkedin.com/in/dongzhanghz

SKILLS

- **Programming Languages**: Python, C/C++, SQL, bash, MATLAB, HTML
- **Machine Learning**: Supervised (linear/non linear/logistic regression, SVM, naive bayes, decision tree, random forest, k-NN), Unsupervised (k-means, GMM, DBSCAN, hierarchical clustering, PCA), Deep learning (CNN, RNN), Natural language processing (SVD, LSA, word2vec/doc2vec), Computer Vision
- **Tools**: Scikit-learn, Tensorflow, Keras, Pandas, NLTK, Gensim, Numpy, Scipy, Matplotlib, Seaborn, Jupyter, Flask, Amazon Web Services, Google Colab, openCV
- **High-Performance Computation**: parallel computing with MPI and OpenMP, job schedulers

EXPERIENCE

- **Data Science Fellow** September 2019 – current
  Insight Data Science — Seattle, WA
  - Consulted for a tech company, and created a tool **Classify3D** to automatically segment and identify objects in 3D images. Each image is **2-10 GB** with **50-100 million** data points (dongzhang84.github.io/classify3d).
  - Used **unsupervised ML** tools (DBSCAN and GMM) to cluster 3D images, and used **computer vision** tool openCV (ORB) to compare image similarities. Identified several classes of objects above **95%** accuracy.
  - Developed the frontend web app using **Python Flask** and HTML.

- **Postdoctoral Researcher**, Computational Astrophysicist September 2018 – August 2019
  University of Michigan — Ann Arbor, MI
  - Developed **high-performance computing** simulations using **half million CPU-hours** to study multiple astrophysical processes in supernova remnants. Generated **∼1 TB** 3D HDF/VTK data from simulations.

- **Postdoctoral Research Associate**, Computational Astrophysicist and Data Engineer
  University of Virginia — Charlottesville, VA September 2015 – August 2018
  - Led two **parallel computing** radiation hydrodynamic simulation projects written in C/C++ using **∼ 2 million CPU-hours** on various supercomputers. Generated **∼10 TB data** for processing.
  - Developed **computer vision** tool and created pipeline in Python to visualize **∼1 TB** multidimensional data generated from simulations. Analyzed data using **Pearson correlation**, **linear/polynomial regression**.
  - **Optimized algorithms** to solve radiative transfer equations to perform the most accurate simulations for many radiation systems, which can be observed by multi-wavelength ground and space telescopes.

- **Graduate Research Assistant**, Theoretical Astrophysicist September 2009 – July 2015
  The Ohio State University — Columbus, OH
  - Built synthetic models using (semi)-analytic methods to explain up-to-date observations of galaxies.
  - Analyzed X-ray data of hundreds of galaxies to constrain key parameters of galactic winds.
  - Led independent projects to develop new models of dark matter structure to explain the origin of early Universe.

DATA SCIENCE PROJECTS

- **Book Recommender System** (NLP modeling and AWS)
  Created content-based book recommender system using users review data scrapped from Goodreads, by NLP word embeddings (TF-IDF, SVD, LSA, doc2vec), and cosine similarity comparison (http://booksrecommend.ml).
- **Flight Delay Predictor** (Supervised Machine Learning)
  Built a machine learning tool to predict flight delays for various airports/airlines, using random forest, kNN, decision tree, support vector machine, and logistic regression (see more on github.com/dongzhang84/Flight_delay).
- **Gas Turbulence Driver** (Computer Vision, Fast Fourier Transform, Statistics)
  Wrote C++ code to generate 3D turbulence in gaseous medium. Created Python pipeline to analyze turbulence data using Gaussian distribution and correlation, and visualize turbulence evolution (movies on youtube).

EDUCATION

- **Ph.D.,** Astrophysics Ohio State University, Columbus, OH, *July, 2015*
- **M.S.,** Astrophysics Nanjing University, Nanjing, China, *June, 2009*
- **B.S.,** Astronomy, *Summa Cum Laude* Nanjing University, Nanjing, China, *June, 2006*