Shu Liao

College Station, TX || (979) 985-1018 || xx9liao@gmail.com || linkedin.com/in/shuliao || github.com/Ikaroshu

EDUCATION

Texas A&M University, College Station, Texas — Sep 2014 - May 2020
PhD in Physics

Georgia Institute of Technology, Atlanta, Georgia — Sep 2018 - Dec 2020
MS in Computer Science

Peking University, Beijing, China — Sep 2010 - Jun 2014
BS in Physics

SKILLS

Programming Languages: Python, C++, SQL, Java
Data Analysis: Numerical simulation, Time series, Bayesian modeling, Machine learning (clustering, regression/classification, neural networks)
Tools & Packages: Numpy, Scipy, Matplotlib, Scikit-learn, Cython, Tensorflow/Keras, AWS

EXPERIENCE

Fellow, Insight Data Science, San Francisco — Jan 2020-May 2020
- Developed a discount forecast model for Steam platform users to find the best time to purchase games
- Scraped and cleaned 30k data from Steam, and utilized random forest to make discount prediction with 60% recall rate
- Launched final product as a web app utilizing AWS Amplify as frontend and Flask on AWS EC2 as backend

Software Developer Intern, OnScale, Atlanta, GA — Nov 2019-Jan 2020
- Developed a kernel smoothing C++ library that is able to smooth 1 million unstructured grid data
- Implemented 3-dimensional sparse matrix representation that improved processing speed by more than 10 times
- Provided Python interface to the C++ library through Cython
- Communicated with project manager effectively to come up with efficient solutions to data post-processing

Research Assistant, Texas A&M University, College Station, TX — Sep 2015-Nov 2019
- Designed Python package PyCEvNS for analyzing experiment data and fitting new physics models using hypothesis testing
- Constructed statistics models for neutrino and dark matter experiment
- Published 9 papers and presented 6 public talks on neutrino phenomenology

PROJECTS

PyCE*ν*NS: python package for cleaning and analyzing neutrino experiment data and fitting new physics models
github.com/Ikaroshu/pyCEvNS
- Optimized for high performance and parallel computing in supercomputer using multinest, numpy and scipy python package
- Implemented data pipelines to clean data and generate posterior probability distributions from models
- Designed package with design patterns for usability and expandability
- Published 6 papers using this package in python by generating results from theoretical models

Tech Blog: a personal blog website built on Django
shusblog.dev/
- Built a data science blog website utilizing Django and Nginx
- Deployed and maintained website on Amazon AWS

Movie recommendation engine development in Apache Spark and Tensorflow
github.com/Ikaroshu/movie_recommendation_engine
- Built data ETL pipeline to analyze 20 million movie ratings and conducted analysis using Spark SQL
- Implemented Alternative Least Square (ALS) model to provide movie recommendations using Spark ML
- Implemented Denoising Autoencoder neural network using Tensorflow to handle cold-start problem