

Shu Liao

San Francisco Bay Area || (979) 985-1018 || xx9liao@gmail.com || linkedin.com/in/shuliao/

SKILLS

Programming Languages: Python, C++

Data Analysis: Numerical simulation, Statistics, Time series, Machine learning, Bayesian modeling, A/B testing

Tools: SQL, Numpy, Scipy, Matplotlib, Tensorflow/Keras, AWS

EXPERIENCE

Data Science Fellow, Insight Data Science, San Francisco

Jan 2020-Present

- Developed a discount prediction forecast for the Steam platform for gamers to find the best time to purchase games
- Scraped and cleaned data from Steam, and utilized random forest model to make discount prediction with 60% recall rate
- Deployed final product as a web app utilizing AWS Amplify and Flask on AWS EC2

Software Developer Intern, OnScale, Atlanta, GA

Nov 2019-Jan 2020

- Communicated with project manager effectively to come up with solutions to data post-processing
- Developed a kernel smoothing C++ library that is able to smooth 1 million unstructured grid data within seconds
- Provided Python interface to the C++ library through Cython

Research Assistant, Texas A&M University, College Station, TX

Sep 2015-Nov 2019

- Designed Python package 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

PyCEvNS: 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 multineest, numpy and scipy python package
- Implemented data pipelines to clean data and generate posterior probability distributions from models
- Implemented various hypothesis tests including Likelihood ratio test, Z-test, and Bayesian test using the Bayes factor
- Published 6 papers using this package in python by generating results from theoretical models

Data Science Blog: a personal blog website built on Django

shusblog.dev/

- Built a data science blog website utilizing Django and Nginx
- Deployed website on Amazon AWS and maintained for the past year

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

EDUCATION

Texas A&M University, College Station, Texas

May 2020(Expected)

PhD in Physics

Georgia Institute of Technology, Online

Aug 2020(Expected)

MS in Computer Science

Peking University, Beijing, China

June 2014

BS in Physics