

FAUSTINE LI

Email: faustinel12@gmail.com

Phone: 678-704-6395

Website: faustinel1.github.io

Education

Duke University, Durham NC

Aug 2016 - May 2018

M.S. Statistical Science, GPA 3.85

California Institute of Technology, Pasadena CA

Sep 2011 - Jun 2015

B.S. Chemical Engineering, GPA 3.44

Work Experience

Statistics Intern, Eli Lilly and Company

May 2017 – Aug 2017

- Used machine learning techniques to classify the severity of retinal disease from images.
- Developed and trained convolutional neural networks (CNNs) using TensorFlow and Keras.
- Diagnosed and resolved model performance issues such as overfitting and unbalanced classes.
- Created a web dashboard that allows users to interactively receive predictions.
- Presented a tutorial company-wide about using deep learning for medical applications.

Research Analyst, Oak Ridge National Laboratory

Jun 2015 – Jul 2016

- Analyzed engine emission data to quantify the pollution impacts of biofuels.
- Performed statistical hypothesis tests such as ANOVA and identified potential outliers.
- Developed tools to automate the processing and identification of SEM images.
- Presented statistical results to researchers and prepared results for publication.

Projects

Bayesian Hidden Markov Models

- Implemented Bayesian inference for HMMs from scratch using Gibbs sampling algorithms.
- Developed a Python package, wrote documentation, and provided examples in Jupyter Notebook.
- Automated testing with Github and CircleCI build tools.

Text Analysis of Job Description

- Collaborated with team members to web-scrape and transform text data from Indeed.com.
- Clustered similar jobs based on descriptions using Latent Dirichlet allocation (LDA).
- Developed a Shiny interface to visualize the results, including an interactive map.

Duke Kaggle Competition

- Placed first in an in-class Kaggle competition, predicting the severity of insurance claims.
- Wrote scripts to reproducibly train and test gradient boosted models using xgboost.
- Used feature engineering, ensembling, and custom objective functions to improve performance.

Skills

- Proficient in Python and R. Currently learning C/C++. Some experience with SQL and Spark.
- Experienced with data science libraries including scikit-learn, pandas, dplyr, and ggplot.
- Other software skills include git, Github, Tableau, LaTeX, Linux, and bash shell.
- Coursework includes machine learning, Bayesian statistics, causal inference, and categorical data.