Frank Rosalia

frank.a.rosalia@gmail.com | 917.613.3514

EDUCATION

LONDON SCHOOL OF ECON.

MSc Data Science
August 2020 | London, England

COLUMBIA UNIVERSITY

BA Mathematics-Statistics
Dec 2018 | New York, NY
Cum. GPA: 3.7
Dean's List (All Semesters)

LINKS

Github:// lse-data-science-student
LinkedIn:// frankrosalia

COURSEWORK

GRADUATE

Artificial Intelligence and Deep Learning
Distributed Computing for Big Data
Bayesian Machine Learning
Machine Learning and Data Mining
Managing and Visualizing Data
Algorithms and Computation

UNDERGRADUATE

Statistical Machine Learning
Elementary Stochastic Processes
Partial Differential Equations
Linear Regression Models
Statistical Inference
Calculus I-IV

SKILLS

PROGRAMMING

Python:
Pandas • Numpy • Scikit-learn • Keras • Matplotlib
R:
ggplot • dplyr • Shiny • tidyr • glmnet
Apache:
Hadoop • Spark • Kafka • Hive
Other:
SQL • Java • Bash

GRADUATE PROJECTS

APPLYING GRAPH NEURAL NETWORKS TO ORGANIZATIONAL NETWORK ANALYSIS

This Capstone project was submitted for my MSc in Data Science at LSE. Using a Microsoft Outlook dataset, I applied state-of-the-art techniques in graph neural networks for organizational network analysis. The research provided major computational speedups and high performance metrics — resulting in a 6000% increase in computation speed against the baseline for network community identification.

BUILDING A TRANSFORMER MODEL TO GENERATE SHAKESPEAREAN SONNETS

Built a Transformer model from scratch using Keras with a Tensorflow backend. The entire corpus of Shakespeare's sonnets was used to train a model to generate sonnets in his style. The resulting sonnets can be read on my GitHub page.

DEVELOPING A DISTRIBUTED A* ROUTING ALGORITHM USING SPARK AND HADOOP

Formulated a graph representation of the United States to run a distributed version of the A^* algorithm using Spark. The algorithm used a subsetting algorithm, creating multiple subgraphs to be distributed across worker nodes for computation. This subsetting process resulted in an average performance speedup of 60% against the baseline.

PREDICTING FATAL MOTOR VEHICLE ACCIDENTS USING THE CRSS DATABASE

Analyzed the CRSS database to develop a prediction model for fatal motor vehicle accidents given demographic characteristics. One of the models resulted in a 98% testing accuracy metric for predicting accident fatalities.

ANALYZING CRIME AROUND THE LONDON UNDERGROUND RAIL SYSTEM

Used webscraping and the folium package (python) to render geographically accurate versions of the London Tube map. This map was then combined with a database of London crime statistics to identify high risk areas within the city for commuters.

EXPERIENCE

THIRD SEVEN ADVISORS | Independent Tech Consultant

February 2020 – Present | New York, NY

- Designed and implemented interactive geographic visualizations to optimize business development and marketing efforts.
- Built automated data collection program to aggregate and grow high-priority clients, ranging from $50-$300 million in assets under management.

PFIZER | Data Analytics Intern

May 2016 – Aug 2017 | New York, NY

- Reduced operating costs by developing a Tableau dashboard, leading company-wide digital asset migration.
- Modernized company knowledge database by transferring information onto a universally accessible platform, giving greater access to all employees.

CERULLO LEARNING ASSISTANCE CENTER | Math Tutor

May 2014 – Aug 2018 | Paramus, NJ

- Successfully assisted at-risk students in completing mathematics courses necessary for graduation through tutoring and supplemental instruction.