

Joseph Pagadora

(909) 260 1642 • joseph.pagadora.c@gmail.com

EDUCATION

University of California, Berkeley | Berkeley, CA

B.A. Statistics and Applied Mathematics

Overall GPA: 3.91

August 2015 - December 2019

Relevant Coursework: Data Structures, Algorithms/CS Theory, Machine Learning, Probability, Statistics, Stochastic Processes, Time Series, Linear Models, Design and Analysis of Experiments, Database Management Systems, Data Science Principles & Techniques, Computer Security

EXPERIENCE AND SELECTED PROJECTS

Database Management System Project | Berkeley, CA

September 2019 – November 2019

- Starting with an existing codebase written in Java, implemented several aspects of a basic database management system, including implementing B+ tree indexes, joins, query optimization, and the lock manager.
- Implemented several join methods using iterator objects: simple/page/index/block-nested loop joins and sort-merge join
- Implemented a basic version of the System R query optimizer: create candidate query plan trees, consider only left-deep plans as a simplifying heuristic, estimate the I/O costs, & use dynamic programming to select estimated best query plan.

Append-Only File-Sharing System Security Project | Berkeley, CA

October 2019

- Designed and implemented a basic file-sharing system that is *append-only* in Golang, with security concerns in mind: stored all data on a third-party server, and stored public keys on another *trusted* key-store server.
- API: initialize user, get user (log in), store/load/append to files, share/receive files, and revoke a user's access to a file.
- Created user objects, initialized with username/password, then created metadata including private/public RSA keys, digital signature keys, and using Google's UUID library, created a UUID for JSON-format storage on the datastore server.
- Files implemented as byte slices in Go: each append by a user is treated as a node of data, encrypted using block-cipher counter (CTR) mode, in a reverse-linked list on the datastore server: this is done in anticipation of revoking access to files.
- Used authentication tools such as HMACs and digital signatures to verify correct user login, file integrity/authenticity, etc.

Intro to Deep Learning (d2l): BART Time Series Analysis Project | Berkeley, CA

July 2019 - August 2019

<https://github.com/jcpagadora/BART-Time-Series>

- Obtained and wrangled BART data (hourly ridership from Berkeley to SF since 2016: wrangled to get daily ridership) and performed a statistical and time series analysis (could see that ridership is generally high in summer, low in winter).
- Fit and analyzed several models: first performed differencing to get rid of the downward trend and the weekly seasonality to try to fit a SARIMA model. Used ACF and PACF plots to find good SARIMA parameters. Best (in terms of AIC) was SARIMA(2,0,4)x(0,1,1,7), including several exogenous variables for holidays, but as there was an additional yearly seasonality, not a good fit overall.
- For forecasting, used a recurrent neural network with keras (LSTM with 3 units with MSE loss). Compared to baseline forecasts (last-value and mean), this had the best performance.

UC Berkeley Statistics | Berkeley, CA

Statistics 134 Undergraduate Student Instructor (uGSI)

August 2018 – December 2018

- Teach and lead a class section for the UC Berkeley introductory probability course in the statistics department
- Held office hours, graded quizzes and exams

“Bear Walk” Map | Berkeley, CA

April 2017

- Implemented the back-end of a map application similar to Google Maps, over the area of Berkeley, CA.
- Used a quadtree of images of the map at various views to implement zooming/searching; utilized a database of highways, roads, and locations to create a graph of locations in Berkeley; implemented A* search algorithm to find shortest path (by distance) between locations; used a trie to enable searching of location names by prefix

SKILLS AND INTERESTS

Programming: Python (including libraries numpy, pandas, scipy, sklearn, etc.), R, Java, Javascript, Linux/UNIX, C, Go

Statistics: Linear Regression, Logistic Regression, Regularization (LASSO & Ridge), Cross Validation, k-Nearest Neighbors, PCA, Hypothesis Testing (including parametric & bootstrap methods, i.e. permutation tests, t, F, chi-square), Decision Trees, Neural Networks, ARIMA models, Monte Carlo simulation, Experimental Design, A/B Testing

Other: Excel, Databases & SQL, Data Visualization (seaborn, Tableau), Docker, Spark, Hadoop, Cloud Computing (AWS), Web scraping, Android Development

AWARDS

- Highest Honors in Mathematics, 2019
- Barry R. and Kang Ling James Scholarship, 2019