Jad Aboul Hosn

linkedin.com/in/jadahosn
jadhosn.github.io

jadhosn@asu.edu
480-417-8889

Education

Arizona State University — Tempe, AZ
MS, Computer Engineering — *May 2019*
Scholarship: *Selected from 15,000 applicants for a full merit scholarship at Arizona State University*

American University of Beirut — Beirut, Lebanon
BE, Computer Engineering, Minor in Business Administration — *June 2017*

Technical Skills

Languages: *Advanced:* Python, R, SQL; *Intermediate:* C/C++, Java, PostgreSQL; *Beginner:* Scala
Machine Learning: Linear and Logistic Regression, LightGBM, SHAP, Bayesian Optimization
Developer Tools: Git, Docker, Azure, Azure Machine Learning, GCP, AWS, PyCharm, Jupyter
Libraries: pandas, NumPy, pySpark, Matplotlib, seaborn, scipy, ggplot2, knitr, tidyr, SHAP, hyper-opt
Related Coursework: Statistical Machine Learning, Artificial Intelligence, Mobile Computing, Distributed Database Systems, Data Visualization, Databases, Object Oriented Programming, Algorithms and Data Structures

Experience

Data Scientist/Engineer — June 2019 – Present
DriveTime — *Tempe, AZ*

Credit Scoring *(Python, Azure Machine Learning, LightGBM, SHAP)*

- Enhanced model performance by 30% compared to legacy leading to an expected $35 million increase in annual profits using advanced ML techniques and Bayesian hyperparameter tuning
- Incorporated state-of-the-art model explainability algorithm (SHAP) to reduce training time from 16 hours to 1 hour by slashing the number of features by 94% while maintaining consistent model performance on unseen data
- Introduced a profit calculator that compares the cost of data products to their predictive lift to rank these products by their net economical impact

Automated Machine Learning Pipelines *(Python, CI/CD, AzureML, AutoML, Docker, Jupyter)*

- Boosted Machine Learning development and deployment velocity by implementing Python pipelines to train, evaluate and deploy ML models using CI/CD and AzureML
- Engineered a parallelizable Python pipeline using Azure's AutoML to explore 60 datasets against 8 ML techniques. Project was featured in Microsoft Build 2020 (video link)
- Built consensus around new technologies and pushed for paradigms that cut the team's Azure consumption by 50% by designing a python template that allows hands-on local debugging before deployment to remote VMs
- Initiated a collaboration channel with AzureML product engineers and industry peers to swap expertise and enhance my team's AzureML system design and development practices

Data Science/Business Intelligence Intern — May 2018 – May 2019
DriveTime — *Tempe, AZ*

Vehicle Depreciation *(R, ggplot2, tidyr, SQL, SQL R Services)*

- Devised and deployed a predictive model in R to rank order vehicles by their expected depreciation (while controlling for the age of the vehicle)

Model Monitoring *(SQL, Informatica, R, knitr, ggplot2, tidyr)*

- Engineered an automated nightly data pipeline that monitors models' efficacy, input/output data drift and generates an automated monthly report for stakeholders

Projects

Peer-to-peer lending model *(Python, flask, Azure, Docker)*

- Trained a classification model (Logistic Regression) to predict the probability of default of p2p loans
- Setup a real-time scoring endpoint using Flask on Azure to ingest, validate and score input data payload