SHREEJAYA BHARATHAN

shreejayabharathan@gmail.com | 916 586 4999 | San Francisco, CA
linkedin.com/in/shreejaya-bharathan | github.com/ShreejayaB | shreejayab.github.io

EDUCATION

MS in Data Science, *University of San Francisco* Jun 2020 (*Expected*)
Relevant Courses: Machine Learning, Deep Learning, Time Series Analysis, Data Ethics, Distributed Computing (Spark), SQL, Python, Design of experiments (A/B Testing), Product Analytics, Data Structures and Algorithms

Bachelor of Technology - Engineering, *National Institute of Technology, Tiruchirappalli* May 2017
Relevant courses: Linear Algebra, Programming, Statistics, Operations Management

WORK EXPERIENCE

Machine Learning Engineer Intern*, Manifold AI, San Francisco* Dec 2019 – present
System identification using deep learning

- Built a module for data simulation in Python which solves differential equations given a set of parameters and time-series data.
- Implemented an end to end machine learning framework to predict the parameters of dynamical systems.
- Built baseline machine learning models such as regularized linear models and Random Forest.
- Applied deep learning to estimate parameters of dynamic systems using CNN architecture(ongoing).

Data Scientist- Decision Analytics*, EXL Services, Bangalore* Aug 2017- Apr 2019
Customer experience analytics for a major energy and utilities company in the US

- Built a multi-class text classification model to categorize customer survey comments using NLP and achieved an accuracy of 0.83.
- Created an ensemble model for predicting complaints using XGBoost and Random Forest algorithms and reduced customer complaints by 20%.
- Worked with the Customer Experience team to identify dissatisfaction drivers from customer surveys and built a user-experience dashboard.
- Developed a churn model after identifying the main cause of customer churn and increased customer retention by 13.8%

Data Analytics Intern, *Ernst & Young LLP, Chennai* May 2016 – Jul 2016
Data visualization for tracking cybersecurity logs

- Created a dashboard using Splunk to monitor real-time event logs, generate alerts and keep track of license usage for the cybersecurity team at Ernst & Young LLP.

PROJECTS

Predicting short-term outcomes in critically ill patients: Predicting *the mean heart rate and arterial pressure* Mar 2020

- Achieved an R2 score of 0.927 using lightGBM, Random Forest and XGBoost models on imbalanced time-series data.

Human Activity Recognition: *Using smartphone and smartwatch data to predict daily activities* Jan 2020.

- Implemented data pipeline in PySpark, leveraging Spark ML, H20, and Auto ML to train models on AWS EMR clusters.

Box office predictions (Kaggle competition): *Predicting the box office revenue of movies* Oct 2019

- Built a scikit-learn pipeline, used Randomized search to perform hyperparameter tuning and ran 5-fold cross-validation to determine the best model and achieved an R2 of 0.85.

TECHNICAL SKILLS

Languages & Technologies: Python, SQL, R, PyTorch, AWS, Docker, MLFlow, git