Hriday Baghar

5219 Brooklyn Ave NE, Seattle, WA – 98105
+1 (206) 229-8614 | hbaghar@uw.edu | hbaghar.github.io

Education

University of Washington ***Seattle, WA***
Master of Science: Data Science ***Mar 2023***

- *GPA:* 3.97/4.0
- *Coursework:* Data Visualization, Statistics and Experiment Design, Software Design, Natural Language Processing, Machine Learning, Time Series, Scalable Data Systems, Deep Learning

Vellore Institute of Technology ***Vellore, India***
Bachelor of Technology: Electronics and Communication Engineering ***May 2019***

- *CGPA:* 8.73/10 (WES evaluation: 3.97/4.0)

Experience

Ocean Data Lab – University of Washington ***Seattle, WA***
Data Scientist (Capstone) ***Oct 2022 – Mar 2023***

- Built an end-to-end machine learning solution on Azure to process distributed acoustic sensing (DAS) data and classify images of acoustic events in the ocean using unsupervised methods to study marine animal behavior
- Worked with 5TB of complex high dimensional data and applied signal processing techniques to prepare the data
- Created low-dimensional embeddings of data using autoencoders in PyTorch and performed GMM clustering of acoustic events

Health Care Service Corporation (Blue Cross and Blue Shield of Illinois) ***Chicago, IL***
Data Science Intern ***Jun 2022 – Aug 2022***

- Contributed to the development of a machine learning model for predicting which diabetic patients are at high risk of developing complications to help member outreach programs to prioritize who to reach out to
- Improved model lift score for a gradient boosted tree (XGBoost) model by 15% (relative) by developing new features and incorporating new data sources such as social determinants of health (SDoH)
- Contributed to data pipeline development in Teradata SQL to leverage various in-house data sources to build datasets to be used for prediction problems in healthcare

Mu Sigma Business Solutions Pvt. Ltd. ***Bengaluru, India***
Trainee Decision Scientist (Data Analyst) ***Jun 2019 – Dec 2020***

- Consulted for the Data Science and Analytics division of a Fortune 500 US Pharmaceutical company on multiple projects for their health economics and clinical trials teams
- Leveraged electronic health record (EHR) data to create a synthetic clinical trial arm for cancer patients saving the client an estimated $5M in recruitment and operational costs. Performed data manipulation, feature engineering, and patient survival analysis using the Kaplan-Meier method in SQL and SAS respectively
- Analyzed and visualized EHR data to understand treatment patterns of cancer patients in the real world in R to identify opportunities to expand drug approval and market share
- Built statistical models like logistic regression and Cox proportional hazards in R to identify factors affecting racial and other disparities in time to treatment access and survival rate for cancer patients

Skills

- *Programming Languages:* SQL (Teradata, Redshift), Python (numpy, pandas, sklearn, PyTorch, Dask), R (dplyr, ggplot2, fable)
- *Statistics and Machine Learning:* Statistical inference (A/B testing, ANOVA), Random Forest, Neural Networks, Regression (linear, logistic, splines)
- *Tools:* Databricks, AWS, Azure, Snowflake, Tableau, Microsoft Excel
- *Soft skills:* Public speaking, Agile practices

Projects

Fantasy Football Analysis [link]***:*** Used Python to extract fantasy statistics from API into an SQLite database and created a UI in Dash to compare player statistics

Visualization of NBA Player Data [link]***:*** Tableau dashboard to analyze and explore NBA player statistics, shooting accuracy and shot distribution over 20 seasons

Movie Review Classification [link]***:*** Used Python to create a text parsing utility to clean movie reviews and prepare a bag-of-words representation and built a logistic regression classifier from scratch to perform sentiment analysis