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***

- Led the development of an end-to-end machine learning solution on Azure to process complex high dimensional distributed acoustic sensing (DAS) data and classify images of acoustic events in the ocean using unsupervised methods
- Reduced dimensionality of data using autoencoders in PyTorch, performed GMM clustering of embeddings to automatically identify 5 main categories of events
- Enabled researchers to analyze occurrence of events from 5TB of originally un-labelled data through visualization of results

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 feature engineering, feature selection, and incorporating new data sources such as social determinants of health (SDoH)
- Reduced effort for 10 different teams by creating reproducible data pipelines in Teradata SQL for features that are being used for prediction problems in healthcare

Mu Sigma Business Solutions Pvt. Ltd. ***Bengaluru, India***
Decision Scientist ***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 an estimated $5M in operational costs. Performed data cleaning, feature engineering, and analyzed patient survival using the Kaplan-Meier method
- Analyzed and visualized EHR data to understand treatment patterns of cancer patients in the real world in R and identified opportunities for drug label expansion in 2 treatment regimens offered by the client
- Developed statistical models like logistic regression and Cox proportional hazards in R to study different patient characteristics and their impact on disparities in time to treatment access and survival rate for patients across 8 different cancer types

Skills

- *Programming Languages:* SQL (Teradata, MySQL, Redshift), Python (numpy, pandas, sklearn, PyTorch, Dask), R
- *Statistics and Machine Learning:* Statistical inference (A/B testing, ANOVA), Random Forest, Neural Networks,
- *Tools:* Databricks, AWS, Azure, Snowflake, Tableau, Microsoft Excel, Git
- *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]***:*** Created 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 in Python