

Hriday Baghar

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

Education

University of Washington
Master of Science: Data Science

Seattle, WA
Mar 2023 (Expected)

- GPA: 4.0/4.0
- *Relevant coursework:* Probability and Statistics, Data Visualization, Applied Statistics and Experimental Design, Software Design for Data Scientists, Natural Language Processing, Machine Learning, Time Series Analysis

Vellore Institute of Technology
Bachelor of Technology: Electronics and Communication Engineering

Vellore, India
May 2019

- CGPA: 8.73/10

Experience

Health Care Service Corporation (Blue Cross and Blue Shield of Illinois)
Data Science Intern

Chicago, IL
Jun 2022 – Aug 2022

- Contributed to data pipeline development in SQL to leverage various in-house data sources to build datasets to be used for prediction problems in healthcare
- Improved model performance by 15% by developing new features and reducing dimensionality of claims data for a model predicting high-risk diabetic patients using feature engineering techniques
- Developed modular and re-usable Python code to evaluate and compare machine learning models through various performance metrics, confidence intervals and SHAP values that was adopted across the organization

Mu Sigma Business Solutions Pvt. Ltd.
Data Analyst

Bengaluru, India
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
- Analyzed and visualized electronic health record (EHR) data to understand treatment patterns of cancer patients in the real world in R to identify opportunities to expand drug market share
- Built statistical models like logistic regression and Cox proportional hazards in SAS to identify factors affecting racial and other disparities in treatment access and survival for cancer patients
- Leveraged EHR data to improve clinical trial recruitment for cancer patients by 20%. Performed data manipulation, feature engineering, and patient survival analysis using the Kaplan-Meier method

Skills

- *Programming Languages:* SQL (Postgres, SQLite, Teradata, Redshift), Python (numpy, pandas, seaborn, sklearn, bs4, pytorch), R (dplyr, ggplot2, fable), SAS
- *Statistics and Machine Learning:* Hypothesis testing (A/B testing, robust statistics, ANOVA), Experiment Design, Survival Analysis, KNN, Decision Trees, Random Forest, K-Means Clustering, Neural Networks (CNN, RNN, LSTM), Regression (linear, logistic, splines, ridge, lasso), Seasonal ARIMA forecasting, SHAP

Projects

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

Analysis of News Article Popularity [\[link\]](#): Performed EDA, hypothesis tests, and linear regression as an inferential tool to understand drivers of article popularity

Forecasting of Air Quality Index Time Series [\[link\]](#): Seasonal ARIMA forecasting models to predict monthly air quality index for 5 different cities using tidyverse tools in R

Statistics for Dummies [\[link\]](#): UI-based EDA and hypothesis testing tool using Python and Streamlit in collaboration with 3 other developers. Applied various software design concepts such as abstraction, inheritance, unit testing and CI

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