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

- *CGPA:* 8.73/10
- *Relevant coursework:* Image Processing, Web Mining, Data Structures and Algorithms, Neural Networks

Experience

Proteus Technologies Pvt. Ltd. — ***Mumbai, India***
Software Engineering Intern — ***May 2021 – Aug 2021***

- Designed an e-learning solution for 200+ students in rural India with limited internet and technology access
- Developed the solution using Node-RED and MySQL, made accessible via a Telegram chatbot

Mu Sigma Business Solutions Pvt. Ltd. — ***Bengaluru, India***
Trainee 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
- Analyzed EHR data to understand treatment patterns of cancer patients in the real world. Visualized and compared treatment switching, and discontinuation of various groups of interest through Sankey Charts in R. Results helped client identify off-label drug usage and analyze market share across different stages of treatment journey
- Built statistical models like logistic regression and Cox proportional hazards in SAS to answer hypotheses about racial and other disparities in treatment access and survival for patients across multiple cancer types. Analysis allowed client to identify areas to focus on in order to improve health equity for groups that are underserved
- Used EHR data to guide decisions on clinical trial recruitment for cancer patients. Performed data manipulation, feature engineering, and patient survival analysis using the Kaplan-Meier method. Client was able to tweak enrollment criteria and improve recruitment by 15% for a clinical trial that was falling short of its goal

Skills

- *Programming Languages:* SQL (Postgres, SQLite), Python (numpy, pandas, seaborn, sklearn, bs4), R (dplyr, ggplot2), 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, subset selection, ridge, lasso), Seasonal ARIMA

Projects

Analysis of News Article Popularity [link] — ***Feb 2022 – Mar 2022***

- Performed EDA, hypothesis tests, and linear regression as an inferential tool to understand drivers of article popularity

Statistics for Dummies [link] — ***Feb 2022 – Mar 2022***

- Built a 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

Movie Review Classification from Scratch [link] — ***Jan 2022***

- Used Python to create a text parsing utility to clean movie reviews and prepare a bag-of-words representation. Built a logistic regression classifier from scratch to perform sentiment analysis and achieved 60% accuracy

Visualization of NBA Player Data [link] — ***Nov 2021***

- Developed a dashboard on Tableau to analyze and explore NBA player statistics, shooting accuracy and shot distribution over 20 seasons