

# Bhairavi Muralidharan

+1 5715233469 | [✉ bhairavi.muralidharan@gmail.com](mailto:bhairavi.muralidharan@gmail.com) | [🏠 bhairavi-m.github.io](https://github.com/bhairavi-m) | [🌐 bhairavi-m](https://www.linkedin.com/in/bhairavi-m) | [📺 bhairavi-muralidharan](https://www.youtube.com/channel/UC...)

## EDUCATION

---

### University Of Pennsylvania

Philadelphia, PA

#### Master of Science in Engineering: Data Science; GPA: 3.85/4.00

Aug 2021 - May 2023

- **Coursework:** Machine Learning, Computational Linguistics, STAT for Data Science, Deep Learning for Data Science, Advanced Data Ethics, Big Data Analytics, Forecasting Methods for Management, Principles of Deep Learning
- **Graduate Teaching Assistant:** Big Data Analytics, Introduction to Software Development, Blockchain and Cryptography

### Christ University

Bangalore, IND

#### Bachelor of Science: Computer Science, Mathematics, Statistics; GPA: 3.85/4.00

Jun 2017 - May 2020

## WORK EXPERIENCE

---

### MongoDB

New York, NY

#### Data Science Intern

Jun 2022 - Aug 2022

- Leveraged high granularity cloud-based server data to identify key patterns in the usage behavior of Atlas Serverless
- Extracted common characteristics from complex workloads to improve CPU resource allocation of Serverless instances using Python
- Constructed a Serverless Workload Characterization model based on clustering techniques that optimized space by 30%

### Computational Social Science Lab - University of Pennsylvania

Philadelphia, PA

#### Graduate Research Assistant

Jan 2022 - May 2022

- Worked on Candidate Coverage for the Penn Media Accountability Project - an interdisciplinary, nonpartisan research project dedicated to enhancing media transparency and accountability at the scale of the entire information ecosystem
- Modelled a Topic Identification classifier to predict whether a news article contains a political candidate as its primary focus
- Quantified the volume of coverage the different candidates received and identified biases in the framing and coverage of these candidates
- Improved model performance by handling class imbalance which increased minority class precision by 5%

### iQGateway

Bangalore, IND

#### Associate Data Scientist

Jul 2020 - Aug 2021

- Conducted research on Automated Machine Learning approaches like pipeline automation and patented multi-metric model evaluation
- Created a modular feature transformation library of an AutoML product that boosts collaboration and improves performance 10 folds
- Optimized MLOps workflow to enable easy transition from model development to inference for production environments

### Hitachi Payment Services

Mumbai, IND

#### Data Analyst Intern

Apr 2018 - May 2018

- Analyzed data on the transactions of ATMs spread across India and came up with a profit/loss tabulation
- Collaborated with the Business Finance team to perform data surgeries using MS Excel that facilitated the automation of timely analysis

## PROJECTS

---

### Lexical Normalization

→GitHub

#### Natural Language Processing, Deep Learning

- Built a Lexical Normalization system that enables efficient information extraction by converting non-standard text to a standard register
- Devised a hybrid model that used a Maximum Frequency Replacement model for In-Vocab words and a Char RNN model for Out-Of-Vocab words while experimenting with different methods of augmenting data, resulting in an increase in the accuracy by 3%

### Session-based Sequential Skip Prediction: Spotify

→GitHub

#### Machine Learning, Evaluating a Recommendation System: Supervised Classification Problem

- Predicted whether users would skip tracks, given their immediately preceding interactions in their listening session
- Preprocessed the data using PCA, compared different approaches like Random Forest Classification with boosting (XGBoost), and extended sequential techniques (pretrained RNN) to examine the effect of the sequence of tracks on the user's behavior

## SKILLS

---

**Programming** Python, SQL, R, Java, C

**Tools/Software** Weights and Biases, Dataiku, MS Office, Git, Atlassian Suite, MongoDB Atlas

**Frameworks** NumPy, Pandas, Scikit-learn, PyTorch, NLTK, SpaCy, PySpark