

A mathematician and a Software Engineer with 7.5 years of work experience, who has passion for innovation and loves solving complex challenges. My career goal is to work as a Machine Learning Engineer.

EDUCATION

- M.S. in Applied Mathematics (Machine Learning); Northeastern University; Fall 2020 – Present; GPA: **3.91/4.0**
- Indian Institute of Technology, Kharagpur – B.Tech. (Hons.) in Civil Engineering; Fall 2009 – Spring 2013.
- MOOC**: MLOps, Deep Learning, Mathematics for Data Science, Data Structures, Algorithms, Object Oriented Design.
- Coursework**: Machine Learning, Computer Vision, Applied Statistics, Linear Algebra, Probability, Mathematical Modeling.

EMPLOYMENT

Senior Software Engineer – Data	Waterline Data Science	Dec 2018 – Aug 2020
<ul style="list-style-type: none">Built a first ever unstructured Data Processor to extract text from images in Lumada Data Catalog (LDC). Developed a novel way to collect training data for improving performance of existing OCR engine by 10 %; acquired at least 2 new clients.Developed an NLP model to resolve user queries by redirecting to FAQs; reducing no. of tickets logged by 60 %. This improved user retention rate of LDC by 150 %; measured with an efficient user visit logging system.Integrated the charting framework ngx-charts into LDC and designed multiple UI components useful for DataOps.Implemented an asynchronous Spark job that helps to sync new/purge ghost content metadata in LDC.		
Software Engineer 2 – Full Stack	Oracle	Aug 2015 – Nov 2018
<ul style="list-style-type: none">Developed RESTful web services for File-Based Data Import (FBDI Oracle) using Java, Spring, Angular, Oracle DB, Docker.Optimized duplicate row detection algorithm using probabilistic approach; reduced time complexity from $O(n^2)$ to $O(n)$.Designed a mobile application to scan OMR sheets of candidates during hiring; reduced time spent in recruitment by 80 %.Expertise in Oracle Business Intelligence Enterprise Edition (OBIEE), BI Publisher (XML Publisher), Accessibility Evaluation.		
Software Engineer – R & D	Altair Engineering	May 2013 – Aug 2015
<ul style="list-style-type: none">Developed a GUI Automation Software by cloning Sikuli. Implemented standard image processing algorithms including Laplace Edge Detection, Pyramid Template Matching, Alpha blending, SIFT descriptor.Adapted Tesseract OCR's code, to increase accuracy in text-recognition for screen fonts from 50 % to 95 %.		

PROJECTS

- Matrix Factorization for User Rating Predictions** Derived update rules and implemented Weighted Alternating Least Squares for predicting missing user ratings of MovieLens data. Improved MSE by 62 % compared to baseline (mean predicting) model.
- Data Modeling using Markov Chain** Performed Time Series Analysis of average runs of opening batters in baseball from 1871 – 2015 with a Markov Chain. Calculated autocorrelation between original time series and a simulated time series. Performed GoF test at 5 % significance level to determine valid states of Markov Chain in a two-step transition matrix.
- FaceNet – Face Recognition** Encoded face image into 128-dimension feature vector (one-shot learning) using FaceNet. Implemented Triplet Loss function to compare Anchor, Positive, and Negative images in training data. Performed face verification and face recognition using the above encodings.
- Debiasing Word Vectors** Used 50-dimensional GloVe vectors to represent words. Performed Word Analogy task and implemented equalization algorithm presented in [Boliukbasi et al., 2016](#) to remove gender bias.
- Northeastern News Updater** Developed a Google Chrome extension to get instant notification updates from NEWS @ Northeastern portal using JavaScript, AJAX, HTML, and CSS. Was awarded a merit scholarship of \$ 25,000.

TECHNICAL SKILLS

- Python, Java, R, C/C++, MATLAB, Mathematica, SQL, PHP, Perl, HTML, CSS, TypeScript, XML, JSON, Visual Basic.
- PyTorch, TensorFlow, OpenCV, NumPy, pandas, Matplotlib, scikit-learn, SymPy, Spark, Hadoop, Kafka, Hive, Zookeeper.
- Git, Jupyter Notebook, Linux, Docker, PyCharm, IntelliJ IDEA, AWS (SageMaker), GCP, Elasticsearch, Angular, Spring, Junit.
- Regression, Classification, Ranking, Recommendation Systems, Clustering, Dimensionality Reduction, Bagging, Boosting, Feature Engineering, Neural Networks, Deep Learning, Computer Vision, Natural Language Processing, Optical Character Recognition.

EXTRA ACADEMIC ACTIVITIES

- Led the Data Club – Spring 2022 at Northeastern. Taught around 100 common interview problems in all levels of difficulty.
- Teaching assistant for Matrix Methods in Data Analysis and Machine Learning (Fall 2021), Calculus 2 (Spring 2021).
- Contributed to an open-source organization named **SymPy** during GSoC-2012 application.