Shashank Bettada

bsshanky@gmail.com | (857) 869-2016 | Boston, MA | linkedin.com/in/shashank-b-s | bsshanky.github.io/Portfolio/

EDUCATION

Master of Science, Data Science — *Sep 2022 - Apr 2024*
Northeastern University, Boston, USA — GPA: 3.81
Bachelor of Engineering, Electronics and Communication — *Aug 2014 - May 2018*
BMS College of Engineering, Bangalore, India — GPA: 8.26

SKILLS

Programming Languages: Python 3, R, Swift, JavaScript, Objective-C
Database Technologies: Relational Databases (MySQL, SQLite3), MongoDB
ML Frameworks and Libraries: PyTorch, Tensorflow, scikit-learn, numpy, pandas, matplotlib, seaborn, ggplot2
Supervised ML Techniques: Regression (Linear, Logistic), Naive Bayes, SVM, Random Forests, CNN (ResNet, VGG16)
Unsupervised ML Techniques: K Means Clustering, Agglomerative Clustering, DBSCAN
Dimensionality Reduction Techniques: PCA, LDA, NMF
Developer Tools: RStudio, Jupyter Notebook, Xcode, VSCode
Cloud Technologies: AWS (EC2, S3, RDS, SNS, Lambda, ELB, ASG, VPC)

PROJECTS

Predicting Breast Cancer Survivability | DS 5110 @ Northeastern University
- Conducted a multifaceted analysis of the METABRIC dataset to **forecast** breast cancer survivability, leveraging diverse analytical models
- Uncovered key insights revealing the Nottingham Prognostic Index (NPI) and tumor size **as crucial and inversely related factors** to survivability
- Employed **four machine learning models,** including SVM, Logistic Regression, Naïve Bayes, and Random Forests, culminating in the Random Forests model reaching a **peak accuracy of 72%** after thorough hyperparameter tuning

Image Classification of CIFAR-10 Dataset | DS 5220 @ Northeastern University
- Developed and compared **two Deep Neural Network models** for CIFAR-10 image classification: **one with and one without convolutional layers**, demonstrating enhanced feature extraction capabilities of CNNs
- Enhanced the classification accuracy from **44% to 60%** using a CNN model compared to the DNN only model, while also achievin**g faster convergence** in the CNN model

XML Parsing and Database Integration | CS 5200 @ Northeastern University
- Engineered a robust and efficient XML parsing algorithm to **process and categorize over 30,000+** medical journal articles, transforming complex and inconsistent data into structured dataframes
- Optimized data extraction code to achieve full document parsing in **fewer than 5 minutes**, and seamlessly integrated the parsed data into MySQL database tables for effective data management and retrieval

Unsupervised Analysis of the LFW Dataset | DS 5230 @ Northeastern University
- Conducted **comprehensive EDA** and preprocessing on the LFW Dataset, including **strategic sampling** for balance, **scaling** pixel values for numerical stability, and applying **PCA** for dimensionality reduction
- **Revealed distinctive clustering patterns**: K Means highlighted variations in facial expressions, Agglomerative clustering effectively grouped distinct features like smiles, and DBSCAN efficiently classified images with similar facial orientations

EXPERIENCE

Software Engineer (iOS), Navigem Data, Bangalore, India — *Oct 2019 - May 2022*
- **Developed** and **delivered** two iOS applications, one for a local client which achieved **1000+ downloads** and a **4.6/5 rating,** and another for an international audience, collectively contributing to **$150,000** in revenue for the organization
- Managed the **system design** phase by decisively advocating for Backend as a Service (BaaS) over a native backend, achieving a **50% reduction** in resources and costs, and ensuring scalability along with feature compatibility to meet user demands effectively
- Implemented various **code optimizations** to efficiently utilize Pay-As-You-Go API services, effectively **reducing overhead costs by 66%** and enhancing system performance

Associate Software Engineer, Accenture, Bangalore, India — *Oct 2018 - Oct 2019*
- **Revamped** 30% of the codebase, strictly adhering to the Single Responsibility Principle, resulting in enhanced testability and maintainability of the software
- Authored **XCTest** suites for **1500+ lines of core application code**, elevating code reliability and minimizing bugs
- Optimized the development process by configuring a Jenkins CI/CD pipeline, automating testing and **ipa f**ile generation, which resulted in a **30% reduction** in testing and deployment time