Bharath Raj Pragada

Boston, MA | +1 (857) 397-2462 | pragada.b@northeastern.edu | linkedin.com/in/bharath-raj | Portfolio | GitHub | Tableau | Medium

EDUCATION

Northeastern University — Boston, MA
Candidate for Master of Science in Data Analytics Engineering, GPA: 4.0/4.0 — Expected May 2025
Relevant Coursework: Statistical Learning for Engineering, Data Mining in Engineering, Computation and Visualization for Analytics, Foundations for Data Analytics Engineering

SASTRA University — Thanjavur, India
Bachelor of Technology - Information Technology (Honors), GPA: 3.5/4.0 — Jul 2022
Relevant Coursework: Machine Learning Techniques, Big Data Analytics, Data Warehousing and Data Mining, Database Management System, Statistical Foundations for Computer Science, Design and Analysis of Algorithms

TECHNICAL SKILLS

Languages: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Tensorflow), SQL, R, C++, Java, HTML, CSS
Analytical Tools: Tableau, MS Office, Jupyter, RStudio, Flourish, Power BI, Google Data Studio, Apache Airflow
Databases: MySQL, MongoDB, Apache Spark, Neo4j, Amazon Web Services (S3, EC2, Route53, Cloud Front, RDS)
Machine Learning: Regression, Classification, Clustering, PCA, Hypothesis Testing, Data Analysis, Decision Modeling
Frameworks: FAST API, Flask, React JS, Java Script (amCharts), JUnit, Pytest, Spring Boot, Django, Spark, Scala

EXPERIENCE

Research Assistant, Northeastern University - Boston, MA — Sep 2023 – Present
- Create and deploy an optimized API utilizing **FastAPI** framework in **Python**, enhancing data retrieval constraints handling
- Research and evaluate machine learning algorithms on experimental Nernst data on model accuracy and scalability
- Executed data loading, cleaning, structuring, and maintenance tasks using MongoDB, to ensure secure data management

Data Scientist, Alfahive - Bangalore, Karnataka — Aug 2022 – Jul 2023
- Created a python script to prepare master data for **CIS CSC** controls, mitigations, MITRE tactics employing **ETL** process
- Designed **incident model** for core risk quantification engine using **Tensorflow** to calculate Annual Loss Expectancy*(ALE)*, aligned to a business function*(stores, infra)* for assessing potential impacts caused by **MITRE ATT&CK** techniques
- Partnered with product team to conduct data aggregations using **MongoDB** into data pipeline to drive insightful analytics
- Revised model by adding a sophisticated filter within layered architecture, resulting in a **80%** reduction in response time
- Collaborated with QA team and improved test performance statistics by **50%** using **Testsigma** and verified UI functionality

Data Science Intern, Alfahive - Bangalore, Karnataka — Feb 2022 – Jul 2022
- Assisted in gathering, cleaning, and re-processing data from different sources, such as databases, APIs, or **web scraping**
- Designed Talent Acquisition Portal using **Spring Boot** and **React JS**, for a Cybersecurity startup enabling job-seekers to apply and helped data team build a robust **FAIR** model that assess, prioritize, and manage information security risks
- Delivered a cutting-edge product with dashboards for data analytics using **amCharts** and Visualization with a team of 3
- Contributed towards development of a predictive model to identify Threat Event Frequency, and Loss Magnitude
- Enhanced user journey, resulting in a **50%** increase in customer value and a **35%** rise in **JUnit** test coverage

Artificial Intelligence Intern, Internship Studio - Pune, Maharastra — Jul 2021 – Aug 2021
- Implemented language translation model using **Seq2Seq** architecture, to convert German to English and vice versa
- Utilized **Keras Tokenizer** to vectorize text data, facilitating effective sequence processing for machine translation
- Streamlined model using **LSTM** layers, trained for **30** epochs with RMS Prop Optimizer, achieving optimal performance

ACADEMIC PROJECTS

EEG Classification Model - *GitHub* — Oct 2023 – Dec 2023
- Employed Recurrence Quantification Analysis *(RQA)* to extract recurrence and determinism features from EEG data
- Trained models such as Logistic Regression, Random Forest, Decision Tree, and XGBoost for multi-class classification
- Achieved superior performance with XGBoost, attaining a **96%** accuracy in binary and **77%** in multi-class classification

Global Exploration of Natural Disasters: Causes, Adverse Effects, and Mitigation Plans - *Google Site* — Sep 2023 – Nov 2023
- Spearheaded a collaborative project to visualize global natural disasters to communicate complex information
- Applied **Datawrapper** and **Flourish**, to create dynamic and engaging visuals, for effective storytelling via **Google Site**

Crime Data Analysis and Visualization - *GitHub* — Sep 2023 – Nov 2023
- Conducted comprehensive exploratory data analysis (**EDA**) and used **Prophet** time series to predict future crime patterns
- Reported on the effects of major events during the dataset period, providing critical insights into shifts in crime rates

Deep Learning Model for Diagnosing Chest X-rays - *GitHub* — Sep 2021 – Dec 2021
- Preprocessed images using Keras ImageDataGenerator and implemented weighted loss to address class imbalance
- Trained a deep learning model and designed a **GUI** for showcasing the predicted disease in chest X-ray using **Python**
- Attained accuracy of **73%** with pre-trained **DenseNet121** (Densely Connected Convolutional Networks) model

Classification and Segmentation of Brain Tumor - *GitHub* — Aug 2021 – Dec 2021
- Implemented a Layered architecture of **ResNet** model to classify the type of brain tumor, achieving an accuracy of **99.2%**
- Constructed a **U-net** segmentation model, which locates the tumor through MRI images, allowing quick diagnosis