

RAGHAV SHARMA

Boston, MA | sharma.rag@northeastern.edu | +1 (781)-298-1547 | [LinkedIn](#)

EDUCATION

Khoury College of Computer Sciences, Northeastern University

Fall 2021 - Present

Master of Science in Data Science

Relevant Coursework: Data management & Processing, Supervised & Unsupervised ML, Deep Learning for Computer Vision **GPA: 4.0 / 4.0**

University School of Information, Communication and Technology –Delhi, IN

2017 - 2021

Bachelor of Technology, Computer Science

Relevant Coursework: Advanced Database management System, Machine Learning, Object Oriented Programming **GPA: 8.60 / 10.0**

TECHNICAL SKILLS

Programming Languages: Python, R, C++, SQL, MATLAB, Apache Spark

Machine Learning: Classification, Regression, Clustering, SVM, PCA, Decision Modeling, Deep Neural Networks, Convolution Neural Network, Computer Vision, Recurrent Neural Networks

Libraries: Numpy, Pandas, Seaborn, Matplotlib, Sklearn, Tidiverse, Dplyr, Tidyverse, Statsmodels, Scipy, Tensorflow, Keras, ggplot2, Pytorch, Spacy, NLTK

Certifications: Machine & Deep Learning, Data Scientist's Toolbox - John Hopkins, Microsoft AI Classroom Certification, Programming essentials in python, Data Structures and Algorithms in C++

PROFESSIONAL EXPERIENCE

Research Internship, Defense Research and Development Organization of India (DRDO), Delhi, India Jan 2020-May 2020

- Developed a Project Report on different models for calculating National Comprehensive Power (NCP), entailing compendious factors like economy, military strength, governance, technology etc
- Visualized various factors affecting the NCP by implementing descriptive models like heat maps, geographic maps, and line graphs
- Calculated **NCP for around 24 leading countries** by executing optimized ML models

Machine Learning (ML) Intern, Indic AI Foundation for Social Good, Delhi, India

May 2019-Dec 2019

- Developed a Python project on Image Scene descriptions for visually impaired individuals
- Flickr8k dataset – **6000 training images, 1000 Validation and Test images** each
- Implemented Text Preprocessing, Image Preprocessing, Train and Test Dictionaries and created Custom Data Generator
- Encoded images using **ResNet50** model with pre-trained imagenet weights, which were fed to the final dense layer
- Partial Captions were fed to embedding layer and then to LSTM model
- Achieved a **BLEU score of 0.45** and **METEOR score of 0.34**

PROJECTS

Time Series Forecasting of HouseHold Power Consumption

Jan 2022-Present

- Analyzed 2,000,000+ rows** of time-series data to forecast the household electricity consumption using Python
- Cleaned, explored and analyzed the dataset to understand trend and seasonality
- Implemented SARIMA, LSTM, Random Forest and Linear Regression to predict energy consumption
- LSTM achieved the lowest **MAPE of 18.7**

Music Genre Classification, Analysis and Audio Generation

Sep 2021-Dec 2021

- Created dataset by extracting various features from multiple audio files reducing **training time for models by 200%**
- Generated images using their spectrograms, which were fed to the CNN model for classification
- Performed in-depth EDA to visualize the audio data using spectrograms, boxplots and co-relation heat maps
- Implemented XGboost, Random Forest, LGBM, Deep Neural Network using Keras and CNN
- Deep Neural Network using Keras gave the **best accuracy (93.37%) and an F1score (0.9013)**
- Generated a playlist of top 10 recommended songs using NearestNeighbours and Cosine Similarity

Project – Reddit Flair Detection

Mar 2020-May 2020

- Data Scrapping from Reddit using PRAW API and performed Exploratory Data Analysis and in depth text preprocessing.
- Deployed an end-to-end webapp on Heroku for real-time data analytics to gain insights and follow trends and identify the flair of the post.
- Implemented flair analysis of various posts leveraging NLP methodologies to extract key features using NLTK, Tf-IDF Vectorizer and TextBlob libraries
- Implemented Machine Learning and Deep Learning models on the dataset to predict the flair of the post from Reddit India.
- Best Accuracy – 89% (Bidirectional LSTMs)**
- Deep Learning Models - 1. Word Embedding + CNN 2.LSTMs 3.Bidirectional LSTMs 4.Hybrid-model CNN-LSTM

COVID19 detection from Chest X-Ray Image

Aug 2020-Dec 2020

- Proposed an effective and convenient way of detecting COVID19 from Chest X-Ray images through deep learning techniques
- Enforced EfficientNet deep learning model on the dataset – Covidx (by far the most comprehensive and largest dataset comprising a total of **13,975 Chest X-Ray images across 13,870 patients**)
- Achieved the best accuracy of **95% followed by CovidNet (93.3%), ResNet50 (90.6%) and VGG-19 (83%)**
- 100% Sensitivity** – Detected every positive patient.
- Presented and published a partially funded **Research Paper at IEEE Healthcom 2020**