ESHA SRIVASTAVA

Boston, MA 02130|667-273-0907|e-mail|LinkedIn|Github

SUMMARY

A highly driven problem solver with experience in handling extensive datasets, implementing cutting-edge solutions, and driving groundbreaking research.

EDUCATION

M.S. in Applied Mathematics(Data Science Track) , Northeastern University **September 2021 - December 2023**
GPA- 3.9/4.0

B.S in Mathematics (Minor in Economics), University of Delhi **July 2017 - November 2020**
GPA- 8.6/10.0

EXPERIENCE

Graduate Research Assistant (DATA Initiative), D'Amore-McKim School of Business, Boston MA **February 2023 – August 2023**

- Scraped the web for 1,00,000+ Amazon camera reviews using Selenium, extracted valuable insights for analysis
- Enhanced the scraper to extract valuable user profile information, providing deeper insights for analysis
- Conducted sentiment analysis of 303 brands over 160 weeks to investigate spillover effects between online and offline domains
- Streamlined data integration using fuzzy matching, thereby enhancing efficiency and accuracy in data analysis
- Delivered actionable insights on online-offline sentiment correlation, thereby empowering stakeholders to prioritize data-driven strategies effectively

Data Analytics & Machine Learning Co-op, Electric Power Research Institute, Palo Alto CA **September 2022 - December 2022**

- Automated outage ticket approval for PJM-ISO, saving 1000+ client hours with a probabilistic classifier using XGBoost
- Extracted meaningful insights from a massive dataset of 1,000,000+ rows and 64 features using advanced data processing techniques and optimized algorithms
- Improved precision and recall for minority class by 11% and 5% respectively by addressing dataset imbalance with class weight adjustments, SMOTE, and TOMEK links
- Developed a Bayesian model achieving 78% accuracy, MAE 23.3, and RMSE 204.8 in predicting overruns, delivered transformative insights to stakeholders, which were showcased prominently in the company's highly anticipated annual research report

Data Scientist Intern, Quantori, Cambridge MA **May 2022 - August 2022**

- Worked with radiologists to develop a novel deep learning solution to quantify edema using AWS S3, EC2 and Sagemaker
- Led data acquisition efforts for securing cloud access to global data sources for the company's project
- Built the data pre-processing pipeline to automate scoring of Pulmonary Edema severity, supporting radiologists
- Leveraged transfer learning and mathematical morphology techniques to perform segmentation on 500,000+ chest x-ray images, using PyTorch and OpenCV
- Ensured adherence to the software development life cycle by following best practices and implemented version control using Git

PROJECTS

Topological Data Analysis (TDA) of Brain Artery Networks(Open Source Contribution): Performed tubular Segmentation on fMRI images of the brain using ITK, built a Machine Learning model that leverages the techniques of persistent homology by taking segmented brain arteries of the control group and Aneurysm patients and investigated correlation between them(ongoing).

Classification and Segmentation of Brain CT Scans: Built a CNN classifier for 700K CT scans using Tensorflow and Keras, Used U NET for Segmentation of brain tumour and used labels provided to assess performance, achieved 0.70 Accuracy on F1 score of 0.6.

Visual Question Answer Model(VQA) using Transformers: Used LXMERT for to answer open ended questions about images in MSCOCO dataset, model aims to enhance navigation for visually impaired individuals and intelligence analysis

Multi-Class Text Classifier for an online e-commerce review: Classified 5,000 reviews into 20 distinct labels. Implemented Naive Bayes, K-Nearest Neighbours (KNN), Decision Tree, and Logistic Regression, to develop separate classifiers. Provided accuracy metrics, detailed classification reports and confusion matrices. This project can be extended for real-time classification of new reviews, showcasing proficiency in text classification and machine learning

TECHNICAL SKILLS

- Python; C++; R; SQL; MATLAB; HTML; Excel; AWS; Docker; MySQL; Git; Jupyter Notebook; PyCharm; Docker
- TensorFlow; keras; PyTorch; scikit-learn; sklearn; SymPy; numpy; pandas; Imblearn; Spark; OpenCV; ITK; Beautiful Soup; Selenium