ESHA SRIVASTAVA

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

SUMMARY

Mathematics grad with expertise in machine learning, data analytics, and large-scale data processing. Proven track record in research, model optimization, and innovative data solutions across diverse projects

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**

- Utilized Selenium, Python, and ProxyMesh to efficiently **scrape over 100,000 Amazon camera reviews**, ensuring comprehensive data collection of both product sentiments and user profiles
- Extracted and processed HTML data elements, refining the granularity and accuracy of the acquired information
- Conducted **deep sentiment analysis across 303 brands over a span of 160 weeks**, revealing crucial spillover effects between online and offline domains
- Implemented fuzzy matching techniques for optimal data integration, significantly boosting the accuracy and efficiency of subsequent data analysis
- **Translated complex data findings into actionable insights** on sentiment correlations, empowering stakeholders to craft and prioritize robust, data-driven strategies

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 100+ client hours** with a probabilistic classifier using XGBoost
- **Extracted insights from 1M+ rows/64-feature dataset** with pandas, advanced data processing, data visualizations, and EDA
- Executed data clustering with K-Means, DBScan, and Gaussian mixture models, refining data segmentation
- Elevated minority class precision by 11% and recall by 5% using imblearn, SMOTE, and TOMEK links to address imbalance
- Developed Bayesian model (78% accuracy, MAE 23.3, RMSE 204.8) for predicting overruns, **prominently showcased in the company's annual 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
- Spearheaded data acquisition, **integrating cloud access to global data sources**, enhancing project's data capabilities
- 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 500K+ 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

Web Scraping Tool: tool for scraping Amazon product reviews and collecting product ASINs from search results. The tool successfully extracts valuable information from Amazon's website for various purposes, such as market research or data analysis

Recommendation System: Used item-based collaborative on segmented customer data on H&M E-commerce dataset, while protecting customer information and improved the quality of recommendations from 0.007 to 0.0204 on MAP metric

Time Series forecasting: A use-case focused tutorial for time series forecasting of Bitcoin price prediction using ARIMA

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

TECHNICAL SKILLS

Languages & Tools: Python, Java, R, SQL, MATLAB, HTML, Excel, AWS S3, MySQL, Git, Jupyter Notebook, PyCharm
Frameworks & Libraries: TensorFlow, Keras, PyTorch, scikit-learn, numpy, SymPy, pandas, Imblearn, Spark, OpenCV, ITK, Beautiful Soup, Selenium
Machine Learning & Data Science: Regression, Classification, Neural Networks, Transformers, Image Segmentation, Topological Data Analysis, Time Series Analysis, ANOVA, A/B Testing