RISHABH SRIVASTAVA

+1 (929) 358-5907 | rs4489@columbia.edu | linkedin.com/in/rishabhsrivastava6 | github.com/RishabhS66

Education

Columbia University — New York, NY
MS in Computer Science (Machine Learning Track) — Expected Dec 2024
Relevant Courses: Natural Language Processing, Machine Learning, High-Performance ML, Databases
TA for: Topics in Software Engineering, Advanced Software Engineering

Indian Institute of Technology Guwahati — Assam, IN
BTech in Electronics and Electrical Engineering, Minor in Computer Science — Jul 2021
Relevant Courses: Computer Vision, Probability, Data Structures and Algorithms
Recipient of Samsung Fellowship Award

Technical Skills

Languages: Python, C++, Java, CUDA, MySQL, MongoDB, MATLAB, React, JavaScript, NodeJS, TypeScript
Technologies/Frameworks: PyTorch, Scikit-learn, TensorFlow, OpenCV, GCP, Kubernetes, Docker, GitHub

Work Experience

Rubicon Robotics, Columbia Build Lab — New York, NY
Software Engineer — Jan 2024 – *Present*
- Responsible for implementing CV algorithms for **SwimBot**, enhancing swimmer detection and performance analysis.
- Utilized **Django** backend to interface with **AWS RDS** hosted on an EC2 instance, coupled with **React** frontend for optimal performance and user engagement.

Adobe Inc. - Adobe Experience Manager (AEM) Assets — Noida, IN
Software Development Engineer Level II — Jul 2021 – Aug 2023
- Leveraged **Sling**, **Apache Jackrabbit Oak** and **OSGi** frameworks to optimize performance and enhance workflow efficiency.
- Spearheaded enhancement of AEM Assets Search by utilizing **Lucene indexing** for efficient information retrieval, Hugging Face's **BLIP** (Bootstrapping Language-Image Pre-training) APIs for asset auto-captioning and **GPT-4** for query pre-processing.
- Led end-to-end implementation of Smart Tags Block-list in AEM Assets Essentials. Users can block smart tags considered inappropriate for an asset and maintain a list of such tags to be blocked for all assets.
- Integrated state-of-the-art **CLIP** (Contrastive Language-Image Pre-Training) based model with AEM Assets' automatic tagging service, improving relevance of generic smart tags by **80%**.
- Improved AEM Assets Essentials' Search Experience using Adobe Firefly in Adobe's Generative AI Hackathon, selected to be presented at **Adobe EMEA Summit 2023**.
- Solved **30+** localization, accessibility and vulnerability issues, fortifying the platform's resilience and reliability.
- Volunteered as the DevOps Champion, managed and maintained the CI/CD pipeline deployed on **Jenkins** to enable seamless integration and delivery of code changes.

Research Experience

Indian Institute of Technology Guwahati — Assam, IN
Undergraduate Researcher under Prof. Tony Jacob — Aug 2020 – Apr 2021
- Proposed a novel architecture called **Unrolled DRAGAN** (Deep Regret Analytic Generative Adversarial Network), skillfully merging unrolling into DRAGAN to mitigate mode collapse.
- Obtained **higher inception scores**, reduction in mode collapse for MNIST, when compared to Vanilla GAN and DRAGAN.

Adobe Inc. — Noida, IN
Media and Data Science Research Intern — Apr 2020 – Jul 2020
- Conceptualised and implemented a new Reinforcement Learning-based algorithm to extract top relevant patterns from temporal, sequential data sets.
- Trained **Deep Q-Network** using **TF-Agents** and extracted patterns ranked by **user-specified measure of interest**.
- Proposed algorithm allowed monitoring and improving user-targeting based on certain Key Performance Indicators.

Hanyang University Ansan, KR

Research Intern May 2019 – Jul 2019

- Designed a new algorithm **Adaptive Shadowed C-Means** (ASCM), to cluster data using **fuzzy** and **shadowed sets**.
- Reduced impact of noise in clustering by keeping outliers concentrated in shadow region.
- Implemented algorithm on **Iris data set** and **Breast Cancer Wisconsin data set**, and demonstrated its use for image segmentation.

Projects

Clustering Emission Intensities Dataset for Better Data Imputation | *Python* | Jan 2024 - Apr 2024

- Implemented clustering techniques on European Central Bank's (ECB) Company Emission Intensities data to facilitate enhanced imputation methods, enabling more accurate predictions in subsequent analyses.
- Successfully employed **TF-IDF** for extracting features from text data, and **PCA** for dimensionality reduction, enhancing computational efficiency and interpretability of the dataset.
- Utilized **DBSCAN** to uncover clusters of varying shapes and sizes, providing valuable insights into underlying structures and relationships within the dataset, crucial for further analysis and prediction tasks.

Lexical Substitution Task with WordNet, Word2Vec Embeddings, and BERT | *Python* | Nov 2023

- Devised a novel fusion strategy, integrating **BERT's** contextual understanding with **Word2Vec's** semantic similarity and **WordNet's** semantic relations, to enhance lexical substitution accuracy and suggest contextually fitting word replacements.
- Attained a precision of **0.189** and recall of **0.189** on 206 attempted instances with mode-specific scoring.

Expense Management Software | *Java, PostgreSQL, React, Redux, Cypress* | Sep 2021

GitHub: RishabhS66/Expense-Management-Software-React-App

- Built a web application to provide an automated solution to record and report business expenses. Features also include approving or rejecting expense claims filed by a user.

Codeforces Problem Recommender | *HTML, CSS, JavaScript* | Aug 2020

GitHub: RishabhS66/Codeforces-Problem-Recommender

- Built a website that analyzes Codeforces users and recommends problems according to their statistics.