SOUNAK RAY

(646) 506–8377 ◇ sr3846@columbia.edu ◇ linkedin.com/sounak-ray ◇ Homepage ◇ GitHub

Education

Columbia University — New York City, NY
M.S. in Computer Science — Sep 2021 – Dec 2022
- Relevant Courses: Database Systems, Engineering Software as a Service, Speech Recognition.

Indian Institute of Technology Guwahati — Guwahati, IN
B. Tech. in Electronics and Communication Engineering, Minor: Computer Science; GPA: 8.44/10.0 — Jul 2016 – Jun 2020
- Relevant Courses: Data Structures and Algorithms, Machine Learning, Computer Vision, Optimization.

Technical Skills

Programming Languages: Python, C++, C, Ruby on Rails, SQL, HTML5/CSS3, MATLAB, Shell Script.
Tools/Frameworks/IDE: Keras, TensorFlow, PyTorch, Git, Jenkins, Docker, Kubernetes, PostgreSQL, Flask, Helm.

Experience

Sprinklr — Gurgaon, IN
Data Scientist, Machine Learning Team — Aug 2020 – Aug 2021
- **Customer Feedback**: Devised efficient and scalable unsupervised learning algorithms to incorporate client feedback in text classification datasets using Python and **improved the F1-Score by 4%.**
- **Insights**: Generated insights from **10,000,000+ social media conversations** with Hugging Face library in PyTorch. Leveraged Multilingual Transformers for Named Entity Recognition **to generate insights across 12 languages**.
- **Deployment**: **Developed and deployed 5 different microservices with REST API endpoints** using Python and Docker, which handled HTTP requests to generate model predictions. **Enabled auto-scaling** using Kubernetes.

Sprinklr — Gurgaon, IN
Data Science Intern, Machine Learning Team — Jun 2020 – Aug 2020
- **Sentiment Classification**: Created a novel embedding model for over 3,500 emojis based on their unicode using Keras. Utilized these emoji embeddings to **improve the F1-score on messages with emojis by 9.7%**.
- **Liveness Verification**: Detected and analyzed over 5,000 videos for the liveness of the person present using Attention ConvLSTM network with TensorFlow. **Achieved an F1-Score of 93% on validation set**.

Projects

Columbia Chat — Current
- Developing a unified messaging platform for Columbia University affiliates using Ruby on Rails (MVC Architecture), HTML and CSS.
- Followed test-driven development by writing unit tests in RSpec and behavior-driven development using Cucumber with **with 92% code coverage**.
- **Implemented email verification** on sign-up with Action Mailer in Rails to ensure only affiliates can join the platform.

Residence Hall Management System (*residence-hall-management.herokuapp.com*) — Sep 2021 – Nov 2021
- Developed an application using PostgreSQL and Flask to manage admissions, finances, dining services of a residence hall.
- Deployed the Flask application integrated with PostgreSQL on Heroku (cloud platform) and wrote RESTful APIs with SQL query statements to perform CRUD operations on the PostgreSQL database.
- **Verified user privileges** for every attempted user action to ensure access to the proper workflow needed by each user.

Online Writer Identification — Jul 2019 – Jun 2020
- Constructed histogram based features using pressure and velocity to represent handwriting characteristics of a writer using Python and **obtained an improvement of 7% recall** on BIT Casia Dataset.
- Encoded temporal information using supervised LSTM autoencoders with Keras. Encoded vectors were classified with SVM using scikit-learn. **Presented a paper (Oral Presentation) published at ICFHR 2020**.

Academic Positions

Graduate Course Assistant for COMS 4995 Elements for Data Science, Columbia University — Fall 2021
Held office hours and recitation sessions to go over relevant concepts and graded coding assignments.

Publications

Sounak Ray, Addrish Roy, Suresh Sundaram. **A Deep Learning Framework with Histogram Features for Online Writer Identification,** International Conference on Frontiers in Handwriting Recognition (ICFHR), 2020. Link to paper