SHANTANU LALITKUMAR JAIN

shantanu.jain@columbia.edu | linkedin.com/in/shantanu-jain/ | shantanu-jain-2142.github.io/ | +1(646)-764-4875
235 W 109th St, Apt 5B, New York, NY 10025

EDUCATION

Columbia University, New York, NY Expected May 2022
M.S. in Computer Science | GPA: 4.11/4.0
Coursework: Machine Learning, Advanced Databases, Big Data & Cloud Computing, Computer Vision, Applied Deep Learning

Savitribai Phule Pune University, Pune, India Jul 2019
B.E. in Computer Engineering | GPA: 3.82/4.0
Coursework: Data Structures, Discrete Mathematics, Object Oriented Programming, Machine Learning, Database Systems

TECHNICAL SKILLS

Programming Languages: Python, C++, C, Java (incl. Android Development), PHP, SQL
Web Technologies: Amazon Web Services, JavaScript, NodeJS
Software Tools/Frameworks: Docker, Tensorflow 2.x, Keras, Pytorch, MongoDB, MySQL, Git, CI/CD

PROFESSIONAL EXPERIENCE

Cohesity Inc., *Software Engineering Intern* – New York, NY Jun 2021 – Sep 2021
- Enhanced Cohesity's SaaS platform (Helios) by enabling expansion and contraction of Cohesity Clusters deployed on multiple cloud platforms including AWS, GCP and Azure.
- Responsible for end to end design of the feature, from design documentation to unit and integration testing.

Siemens PLM Software, *Software Engineer Associate* - Pune, India Jul 2019 - Dec 2020
- Deployed and maintained the backend infrastructure on AWS as part of DevOps in the product research team.
- Developed a serverless architecture for automated maintenance of the entire MongoDB infrastructure.
- Migrated the infrastructure to various AWS regions using Terraform with Terragrunt.

7Targets, *Software Intern* - Pune, India Apr 2019 - Jun 2019
- Implemented an AI assistant for conversing and maintaining hot leads and saving up to 60% time for sales team.
- Utilized pivot-based machine learning for data augmentation, achieving an increase of 125% of original data.

Anomaly Solutions Pvt. Ltd., *Software Intern* - Pune, India Feb 2018 - Apr 2018
- Developed an android application for classifying different plant species based on images of leaves, flowers and stem.
- Applied convolutional neural network (CNN) for multi-organ plant classification with an accuracy of ~81%.

ACADEMIC PROJECTS

Camelyon-16 Challenge: Identification of Cancerous Cells Mar 2021 - Apr 2021
- Worked on a dataset of 21 gigapixel pathology images, to identify cancerous tissues in the slides.
- Applied data augmentation techniques with multiple zoom levels of the slides for reducing overfitting.
- Used fine-tuned transfer learning with 2-tower InceptionV3, to achieve an accuracy of 96%.

Information Extraction from Unstructured Web Database Feb 2021 - Mar 2021
- Created an information extraction system, for fetching structured tuples from unstructured documents on the web.
- Utilized iterative set expansion algorithm, by fetching a seed tuple to begin the process.
- Used spacy for named entity tagging, and SpanBERT for identification of relations from tokenized sentences.

Voice Controlled Photo Album Feb 2021 - Mar 2021
- Created a web application hosted on S3, for uploading and searching photos using text and voice inputs.
- Quick search access through elasticsearch, with authentication and authorization configured through AWS Cognito.
- Optimized the workflow, by using AWS Codepipeline for continuous integration and deployment.

E-Voting using Blockchain Jul 2019 - Sep 2019
- Designed a scalable and secure online voting system based on blockchain and smart contracts in Python.
- Synchronized blockchain across states with a combination of Merkle trees and proof-of-work consensus algorithm.
- Presented blockchain voting system as part of a peer-to-peer network of miners and users.

Comprehensive Developer Assistant (CODA) Aug 2018 - Mar 2019
- Built a chatbot assistant for aiding developers by providing integration with VCS and terminal commands in Python.
- Generated automatic summary of a given code snippet (SQL statements) with a BLEU-4 score of 19.2 with +1 smoothing.
- Applied LSTM seq-to-seq TensorFlow model for generating SQL query from previous preceding keywords.