

Vineet Srivastava

Email: vsriva4@uic.edu | Mobile: (312)-709-1929 | Address: South Loomis Street, Chicago, IL-60607

Portfolio/Website: <https://vineet1409.github.io/> | Patent: <https://patents.google.com/patent/US20200039784A1/>

LinkedIn: <https://www.linkedin.com/in/vineet-srivastava-452730165/> | GitHub: https://github.com/vineet1409/Data_Science_Projects/

Experienced Software professional with over **5 years** of Industrial experience in **Machine Learning, Big Data Engineering/Cloud Analytics, and IoT** across diverse domains, including **Healthcare, Semiconductor, and IoT**. Skilled in deducing meaningful insights from unknown data and making data-driven decisions. Proven track record of delivering high-quality projects within tight deadlines and budgets.

Skills

- **Programming:** Python, R, Embedded C, OOPs using Python, Linux System Programming, Multithreading
- **Data Structures:** Algorithm Complexity, Arrays/Lists, Strings, Linked-Lists, Trees, Graphs, Stacks & Queues
- **Exploratory Data Analysis:** Pandas, NumPy, Seaborn, Matplotlib, Data Cleaning, PyTorch, Tensor-Flow, Feature Engineering, familiar working & performing analytics on different file formats- CSV, Parquet, Excel, JSON, YAML, Config.ini, etc.
- **Statistics & Machine Learning:** Hypothesis & Chi-Square Testing for feature selection, Central Limit Theorem, ANOVA, T-Test, Linear & Logistic Regression, Gradient Descent, Regularization- Ridge and Lasso, Cross Validation, Hyperparameter Tuning, Decision Tree, Ensemble Techniques- Bagging & Boosting (Random Forest, AdaBoost, XgBoost), KNN, K-Means, Hierarchical Clustering, DBSCAN, Data Drift and Data Leakage
- **Big Data Analytics & Cloud:** SQL with Azure, Oracle/MySQL, PySpark, AWS – EC2, IAM, S3, Lambda, AWS Glue, Athena, Azure Synapse Analytics, SQL Pool, Kafka, NoSQL- MongoDB, ETL pipelining, Data Lake, CI/CD pipelining using GitHub Actions, Terraform, Airflow, Kinesis
- **Natural Language Processing (NLP):** NLTK, Spacy, TensorFlow, Text Processing- Tokenization, Stopwords, Stemming/Lemmatization, Bag-of-words, TF-IDF, Word Embedding with Gensim- word2vec, fasttext, RNN, LSTM
- **Deep Learning and Computer Vision:** Artificial Neural network (ANN), CNN- Image Classification, Computer Vision- Objection detection and Tracking, OpenCV, Scikit-image, Face-Embedding (MTCNN, FaceNet), Image Segmentation (VGG, Unet++)
- **Tools & Process:** Appium for Mobile App Automation, Docker, GitHub, Web-Scrapping (REST-APIs, BeautifulSoup), Android (ADB) Debugging, Agile Process, Kanban, Power-BI, Azure ArcGIS for geospatial mapping, AWS QuickSight, Plotly, Dash

Industrial Experience

Graduate Research Assistant/Graduate Hourly
UI Health (UIC School of Public Health/UIC Cancer Center)
09/2022- Present

Project: Designed an end-to-end big data and ML pipeline for [construction of synthetic syndromic surveillance architecture on Azure Cloud](#)
Tech-Stack: Python, PySpark, Azure Synapse Analytics, Azure Spark SQL, Azure Databricks, Machine Learning, Azure Data factory, Data Lake

- Analyzed the Illinois in/out-patient data acquired by **COMPdata Informatics** and statistically matched patient records with **RTI US Synthetic Household population database** to get location/coordinates and visualize on ArcGIS using geospatial mapping.
- Analyzed and streamlined the cumulative data (**in Tbs**) in **Azure Synapse Analytics** using **Python (Azure PySpark)**. The data generated was stored into **Azure Data Lake** and after filtering with ICD-10 codes was ingested into Azure Databricks for predictive classification-based ML models creation and pipeline.
- Represented the team by displaying the project's architecture and design at a poster event at [2022 Institute for Public Health and Medicine Population Health Forum](#) and **10th anniversary of Northwestern University**, nominated as a **finalist for the Rowland "Bing" Research Award**.

Project: To Segment the Polyps from Colonoscopy Images using Deep Learning Image Segmentation Techniques

Tech-Stack: Python, Pytorch, OpenCV, Pandas, Numpy, Scikit-learn, VGG architecture, Unet++ network, Computer Vision, CNN, Image Augmentation

- **UI Health CVC-trials Dataset:** Database consists of frames extracted from **colonoscopy** videos. The dataset contains several examples of polyp frames & corresponding ground truth for them. The Ground Truth image consists of a mask corresponding to the region covered by the polyp in the image. The data is available in both .png and .tiff formats.
- Performed **data augmentation using pytorch** and identified the key metrics for model creation.
- Used the **VGG and Unet++** architecture to train and build model for polyps segmentation.

Senior Engineer
Qualcomm India Private Ltd.
11/2021- 07/2022

Project: Developed an end-to-end pipeline to [derive critical metrics and create predictive models to identify Bluetooth connection failures](#)
Tech-Stack: Python, PySpark, Machine Learning, Data Lake Architecture, AWS-S3, IAM, Lambda, Glue-Studio, Sagemaker, Athena, QuickSight

- **Orchestrated entire pipeline-** data collection, data storage, data transformation and analytics with **AWS Glue-PySpark and Lambda**, visualization of important metrics on QuickSight and created **predictive classification-based models** to identify connection failure patterns.
- Developed pipeline helped production team to identify faulty wearable devices and **improved efficiency of production by 22%**

Senior Software Engineer
Mirafra Technologies India Private Ltd.
11/2021- 07/2022

Project: Developed an end-to-end pipeline for [Face Authenticator System with deployment on Azure](#) for Healthcare Industry (Security).

Tech-Stack: Python, Face-Embedding (MTCNN, FaceNet), MongoDB, Terraform, Azure App Services, Azure Container Registry, CI/CD pipeline.

- Designed and developed a modern Face Authentication System with **deployment on Azure** which includes state-of-art algorithms like **MTCNN, FaceNet** to detect face and generate face embedding for a healthcare domain-based client.
- Developed system had **endpoints which can be integrated to any device**, and this helped client in meeting their multi-faceted healthcare requirements from **patient identification, access control to integration with Electronic Health Records (EHRs)**.

Project: Designed an end-to-end ML pipeline to [predict the battery status of IOT Sensor as part of smart building automation](#)

Tech-Stack: Python, Zigbee, Kafka, MongoDB, Machine Learning, AWS-EC2, S3, ECR, Dockers, CI/CD pipeline using GitHub-Actions

- Streamlined complete project and managed a team of 4 junior engineers from **data ingestion, validation, transformation, model creation to deployment** by CI/CD pipelining using **GitHub actions to AWS EC2**. The automated pipeline helped **reduce cost of installation by 30%**

Research/Patent

- Working on the healthcare-research as part of the [synthetic syndromic surveillance project](#).
- US Patent: [Detecting Elevator Mechanics inside Elevator Systems.-](#) RSSI data analysis & creation of fine-tuned Trilateration Algorithm.

Certifications/Awards

- HIPAA Associate (HIPAAA)- issued by KnowBee
- Group 1. HSP, Biomedical Research Investigators and Key Personnel
- Azure Data Studio- issued by LinkedIn Learning
- Employee of the Month Award- issued by L&T Technology Services

Education

- | | |
|--|-----------------------------|
| • Master's in Business Analytics, University of Illinois at Chicago - USA, CGPA: 3.78/4.0 | 08/2022- 12/2023 (Expected) |
| • B-Tech in ECE, VIT Chennai - India, CGPA: 8.16/10.0 | 05/2012- 05/2016 |