VINEET SRIVASTAVA

Email: vsriva4@uic.edu, srivineet93@gmail.com | **Mobile**: (312)-709-1929 | **Address**: South Loomis, Chicago, IL, 60607
Portfolio/Website | LinkedIn | GitHub | Healthcare-Analytics Research | Patent | Blogs

*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, Finance/Risk Analytics, and IoT.** Skilled in deducing meaningful insights from unknown data and make data-driven decisions. Delivered more than **25 high-quality projects** within tight deadlines & budget.*

SKILLS

- **Statistics & Machine Learning (ML)**: 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, CatBoost), KNN, K-Means, DBSCAN, Recommendation System ML Design, Data Drift and Data Leakage
- **Data Analysis & Visualization:** Exploratory Data Analysis: Pandas, NumPy, Seaborn, Matplotlib, Tidyverse, Sklearn, PyTorch, Tensor-Flow, Feature Engineering, Power-BI, Tableau, AWS QuickSight, Analytics on different file formats- CSV, Parquet, Excel, JSON, YAML, Config.ini
- **Programming Languages**: Python, SQL, T-SQL, R Programming, Embedded C, OOPs, Data Structures & Algorithms
- **Big Data Analytics, Cloud & MLOPs:** SQL with Azure, Oracle/MySQL, PySpark, AWS - EC2, IAM, S3, Lambda, AWS Glue, Athena, Sagemaker, Azure - Azure Synapse Analytics, Azure Blob Storage, Databricks, Azure ML, Azure Web Apps, Kafka, NoSQL- MongoDB, ETL pipelining, Data and Delta Lake Solutions, CI/CD pipelining using GitHub Actions, Terraform, Airflow, Kinesis, Jenkins
- **Natural Language Processing (NLP):** NLTK, Spacy, Text Processing- Tokenization, Stopwords, Stemming/Lemmatization, Bag-of-words, TF-IDF, Word Embedding with Gensim- word2vec, fasttext, RNN, LSTM, GRU, Bi-LSTM, Transformers, BERT (ClinicalBERT, BioBERT), BART
- **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), DeepFace, Image Segmentation (VGG, Unet++)
- **IOT:** Bluetooth Low Energy (BLE), Zigbee, UWB, TCP-IP, MQTT
- **Tools & Process:** Appium, Docker, GitHub, Web-Scrapping (REST-APIs, Beautiful-Soup, Selenium), Android (ADB) Debugging, Agile Process, Kanban, Anaconda, Jupyter Notebook, Google Colab, Visual-Studio, Pycharm, Wireless Sniffers- Wireshark, Ellisys

EXPERIENCE

Graduate Research Assistant (Big Data and Machine Learning) – UIC School of Public Health (Healthcare Domain) 08/2022 – Current
Project: Designed an end-to-end Big Data and ML pipeline for construction of synthetic syndromic surveillance architecture on Azure & AWS Cloud.
(Python, SQL, PySpark, Azure, AWS, Machine Learning- Classification and Clustering)

- 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.
- Streamlined and orchestrated entire dataset pipeline by creating data lake solutions in Azure (Azure Synapse) and AWS (AWS Glue and Sagemaker), created Classification models for different disease classification like Covid-19, Lung Cancer, Opioids overdose, etc. based on ICD-10 code filtering in Sagemaker with deployment as endpoint. Developed pipelines to monitor and retrain the model, if required, based on data-drift and data-leakage.
- Created Clustering model using K-means, DBSCAN and Gaussian Mix Models to analyze and visualize different regions around Cook-County Chicago region affected by Opioid-overdose, Homicides (Gun-Violence) and Covid Deaths based on Geographical coordinates and Zip-Codes on ArcGIS and AWS Geospatial Platform.
- Represented 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.

Senior Engineer- Qualcomm India (Semiconductor/IOT Domain) 11/2021 - 07/2022
Project: Developed an end-to-end pipeline to derive critical metrics and create predictive models to identify Bluetooth connection failures
(Python, PySpark, Machine Learning, Data Lake Architecture, AWS-S3, IAM, Lambda, Glue, Sagemaker, Athena, QuickSight, CI/CD Pipeline)

- 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 workflow helped production team to identify 1000+ faulty wearable devices and improved efficiency of production by 22%.

Consultant- Capgemini India (Healthcare IOT Domain) 02/2021 -10/2021
Project: Developed an end-to-end pipeline for face authenticator system with deployment on Azure for Healthcare Industry (Security).
(Python, PySpark, Face-Embedding- MTCNN, FaceNet, MongoDB, Terraform, Azure App Services, Azure Container Registry, CI/CD pipeline)

- Designed 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. Collaborated with Product Owners to convey non-technical information to stakeholders.
- Implemented architecture and 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 EHRs. This ensured an increase in client satisfaction rating from 70% to 92%.

Senior Software Engineer- Mirafra Technologies India (IOT-Smart Building Automation) 10/2019 - 02/2021
Project: Designed an end-to-end ML pipeline to predict the battery status of IOT sensor as part of smart building automation.
(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 EC2. The automated pipeline helped to reduce the cost of installation of sensors for building automation by 30%.
- Employed Agile techniques such as user story mapping, backlog refinement, and iterative development to deliver high-quality software within tight deadlines. Fostered a collaborative and adaptive work environment, resulting in increased team productivity and customer satisfaction.

ACHIEVEMENTS

- Healthcare Big Data Research Work/Paper | US-Patent Published on Wireless Data Analytics & Indoor Localization | HIPAA Associate and CITI (HSP, Bio-Medical) Certifications | Azure Data Studio LinkedIn Learning Certification.
- Employee of the Month Award by L&T Technology Services, India | Led the Robotics Team at IISC Bengaluru, India | AWS Blog Post for Social Determinant of Health Research (In Progress) in collaboration with UI Health.

EDUCATION

- **University of Illinois Chicago, USA | Master of Science in Business Analytics** (**CGPA**- 3.9/4.0) 08/2022 - 12/2023 (Expected)
- **VIT University, INDIA | Bachelor of Electronics & Communication Engineering** (**CGPA-** 3.6/4.0) 05/2012 - 05/2016