VINEET SRIVASTAVA

Data Analysis & Cleaning ► Machine Learning ► Big Data Analytics ► IOT & Wireless Communication

vsriva4@uic.edu | +1 (312)-709-1929 | LinkedIn | GitHub | Portfolio | Patents | Healthcare-Research | South Loomis Chicago,IL-60607

Professional Experience

- **Qualcomm Bengaluru, India-** *Senior Engineer* (11/2021- 07/2022)
  Technologies/Tech-Stack: Bluetooth Low Energy(BLE) and LE Audio (Basic Audio Profile-BAP, Common Audio Profile- CAP, GATT, Enhanced GATT Caching, BLE Connections), Python, Embedded C, **Wireless Data Packets Analysis, LE Audio Profiles Automation Framework Development**, **Machine Learning & Predictive modelling**
- **Capgemini Pune, India-** *Consultant* (**Client: Dexcom San Diego, USA**) (02/2021- 10/2021)
  Technologies/Tech-Stack: Continuous Glucose Monitoring System (CGMS) using BLE**, IOT Healthcare Analytics**, Python, EDA & Feature Engineering, BLE packets debugging & Automation.
- **Mirafra Hyderabad, India-** *Senior Software Engineer* (10/2019- 02/2021)
  Technologies/Tech-Stack: Smart Building Automation, Zigbee, Data Analysis and Machine Learning, Python, Data pipelining and Cloud Deployments, Object Tracking using Computer Vision
- **L&T Technology Services, India-** *Engineer* (09/2016- 02/2019)
  Technologies/Tech-Stack: Indoor Localization & People Tracking using BLE, Predictive Machine learning models- such as KNN, Linear Regression, BLE RSSI Filtering- Kalman Filter, Moving Average, Published Patent on designing a fine-tuned Trilateration Algorithm for indoor localization, Python, Embedded C

Industrial Projects

a. Title: ***Deriving Critical Metrices & Create Predictive Model to identify Bluetooth Connection Failures***

***Qualcomm India**, Senior Engineer, Project-Tenure: (02/2022- 06/2022)*

Problem Statement:

- The production line team was facing regular **BLE connection failures** between wearable devices like **Fitbit and BLE receiver(Phone)** thereby leading to loss of real time sensor data (BLE packets). The team wanted to identify the important metrices and understand the pattern behind connection loss.
- The task was to **create a pipeline to analyze the metrices from sniffed BLE packets and perform quick analytics. A predictive classification-based machine learning model was to be designed** to predict the connection failures based on identified critical parameters during packet exchange between BLE Transmitter (Wearable device) and Receiver (Phone).

Tech-Stack/Tools: Bluetooth Low Energy (BLE), BLE packets debugging, Python, PySpark, AWS- S3, Glue, Glue Studio, Athena SQL Queries, QuickSight, ML Models- Logistic Regression, Decision Tree, Random Forest, AdaBoost, XgBoost.

Roles & Contribution:

- **Data Pipelining & Analytics**

✓Ingested **sniffed CSV files into S3 Raw data bucket** as part of the **data lake** architecture. Data was catalogued using **AWS Glue.**
✓Created an AWS Glue job to read the raw CSV files from S3-raw bucket, convert to **parquet format** and ingest in **S3-Clean Bucket.**
✓Created another glue job to **read the parquet files from s3-Clean Bucket and perform exploratory data analysis and feature engineering using Python** to derive **critical metrices behind connection failures** and store the same in data lake **(S3-Analytics bucket)** and create a database for the same for **Athena queries**.
✓Used AWS Athena Services to query the database of BLE metrices created as part of AWS Glue jobs to be used by AWS **QuickSight** dashboard for analytics and visualization.

- **Machine Learning (ML) Pipeline**

✓Developed a **classification-based ML** algorithm by reading the data from AWS-S3 Analytics bucket created as part of AWS glue jobs to predict the connection failure/signal loss based on critical parameters identified such as BLE time to connect, Time in connection, Time to disconnect, Encryption validation, ACK of sensor-data.
✓The **Adaboost performed better** when compared to other models with **accuracy of 66% and recall of 72%.** The trained model along with other artifacts was stored in **AWS-S3 artifacts model**.

Results: The entire pipeline, analytics on Dashboard along with trained model**, helped the production line team quickly to identify the faulty wearable devices or the ones performing poor due to frequent connection failures**

b. Title: ***Predicting Lifecycle of IOT Sensor by designing an end-to-end ML Pipeline (Building Automation)***

***Mirafra Technologies India**, Client: Honeywell, Senior Software Engineer, Project-Tenure: (04/2020- 01/2021)*

Problem Statement:

- As part of the **smart building automation, Zigbee-based** end-devices such as **THL (temperature, humidity, and light) sensors** are used to automatically adjust room temperature or light based on their readings. These sensors were, however, subjected to **battery drain**, and hence their lifecycle was compromised.
- The team was endowed with the responsibility of **creating an end-to-end ML pipeline** with cloud deployment to predict the sensor's lifecycle (**whether endangered or working normally**), thereby **reducing the cost of frequent setup and installation**.

Tech-Stack/Tools: ZigBee, Python, Confluent-Kafka, NoSQL-MongoDB, Docker, EDA & Feature Engineering, Machine Learning-Classification based Algorithms, MQTT, GitHub Actions, AWS- EC2, ECR, S3, CI/CD Pipeline, Deployments, Kanban, Jira, VS-code

Roles & Contribution:

- **Data Pipelining & Engineering**

✓ Developed a **python-based Gateway Application** running on Raspberry-Pi to receive the readings from THL (Temperature, Humidity & Lux) Sensor along with Target Feature (Sensor-Life Endangered:1, Working Normally: 0), serialize & preprocess the data and send to **Confluent-Kafka Topic** and **MQTT server for actions.**

✓ Developed a Consumer application on a Linux Machine to consume the available data from **Kafka-Topic** and send the same to **MongoDB NoSQL** database cluster to be used for analytics & predictive modelling.

- **Machine Learning (ML) Pipeline**

✓ Developed and actively supported by team to create ML components by consuming data from MongoDB- **Data Ingestion, Data Validation, Data Transformation, Model Creation, Model Evaluation and Model Pusher.**

✓ Developed Classification based ML algorithms like **Decision Tree, Random Forest, AdaBoost & Logistic Regression** with EDA, and Feature Engineering. **Handled data imbalance** using techniques such as **SMOTE and Oversampling**.

✓ Discussed with **Technical Architects and Product owners** to counter issues of **Data Leakage and Data Drift** during data validation phase.

✓ Evaluated the model using different metrices such **Precision, ROC-Curve, Recall & F-Score** and selected the best model.

✓ **The Random Forest model** gave better accuracy and precision, **close to 80%** and helped in proper identification of faulty sensors. The hardware team was quick to fix those sensors in case, anything was wrong with their controllers.

- **CI/CD Pipelining for DOCKER image deployment on AWS EC2(self-hosted runner) using GitHub Actions**

Results: The entire pipeline developed helped to reduce the burden of frequent setup. **The installation costs went down by almost 30%.**

c. Title: *Construction and Design of Synthetic Syndromic Surveillance Cloud Architecture-Azure*

UIC School of Public Health, *Graduate Research Assistant, Project-Tenure: (10/2022- Present)*

Tech-Stack/Tools: Python, PySpark, Azure Synapse Analytics, ETL pipelining, Microsoft SQL, Data Cleaning and Feature Engineering, Machine Learning

- Analyzed the **Illinois in/out-patient data acquired by COMPdata informatics** and statistically matched the patient records with household population records from the **RTI US Synthetic Household population database** to generate synthetic population household data.
- Analyzed 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 an **Azure SQL Server**.

Skills & Technologies

- **Programming: Python,** R, C# , 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, **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, Naive Bayes, Data Drift and Data Leakage concepts.
- **Natural Language Processing & Deep Learning:** NLTK, Spacy, Tensor-Flow, Gensim, **Text Processing-** Tokenization, Stopwords, Stemming/Lemmatization, Bag-of-words, TF-IDF, **Word Embedding-** word2vec, fasttext, RNN, LSTM, NER, **Artificial Neural network (ANN), CNN**- Image Classification, **Computer Vision-** Objection detection &Tracking, OpenCV, **NLP on Clinical Trial Datasets to develop keyword search**.
- **Big Data Analytics & Cloud: SQL** with Azure, Oracle/MySQL, **PySpark, AWS** – EC2, IAM, S3, Lambda, AWS Glue, Athena, **Azure** Synapse Analytics, SQL Pool**, Kafka, Airflow, NoSQL-** MongoDB, **ETL** pipelining, **CI/CD pipelining** using **GitHub Actions**
- **Visualization Dashboards**- **Power-BI**, Azure ArcGIS for geospatial mapping, AWS QuickSight, Plotly, Dash
- **Tools & Process: Appium** for Mobile App Automation, **Docker, GitHub**, Web-Scrapping (REST-APIs, Beautiful-Soup), Android (ADB) Debugging, **Agile** Process, **Kanban**, Testing Frameworks- PyTest and Unittest, **HIPAA** compliant.

Research/Patient

- Working on the healthcare-research as part of the synthetic syndromic surveillance project, presented our research work poster at 2022 Institute for Public Health and Medicine Population Health Forum at **Northwestern University**, nominated as a **finalist** for the **Rowland "Bing" Chang excellence in Research Award.**
- US Patent: Detecting Elevator Mechanics inside Elevator Systems.- RSI data analysis & creation of fine-tuned Trilateration Algorithm.

Education

- **Master's in Business Analytics**, University of Illinois at Chiago- USA, CGPA: 3.78/4.0 **Expected December 2023**
- **B. Tech in Electronics & Communication**, VIT Chennai- India, CGPA: 8.16/10 **May 2012-May 2016**