Sharvari Kalgutkar

323-470-3032 San Jose, CA sharvari.kalgutkar82000@gmail.com sharvarikalgutkar GitHub Portfolio

Education

Masters in Data Science, University of Southern California **Aug. 2022 – May 2024**
Fairness in AI, Machine Learning, Deep Learning, Data Mining, Data Analytics, Database Systems *(CGPA 3.87/4)*

Bachelors in Electronics Engineering, Sardar Patel Institute of Technology **Aug. 2018 – May 2022**
Data Structure, Algorithms, Statistical Analysis, Object Oriented, Marketing, Management *(CGPA 9.52/10)*

Experience (1+Year)

Data Scientist, *University of Southern California* **Feb 2024 – April 2024**
- Evaluated **10 Deep Learning** models using **TensorFlow** on **Image data** leading to **45%** cost improvement.
- Researched **Neural Network** performance on non-identical data using **distributed computing** like FedAvg.
- Awarded **Best Data Science Team** for outstanding **interdisciplinary** data science research by **USC CKIDS**.

AI Engineer, *Scientist Technologies* **Nov 2021 – May 2022**
- Devised **Computer Vision** Road quality tracking system using models like Faster R-CNN, yielding **84% precision**.
- Optimized YOLOv5 **Object Detection** model for diverse vehicle classification, resulting in **64% mAP**.
- Cross-collaborated on **5 Python algorithms** for road safety data analysis using **Agile**, achieving a **91% R2 score**.
- Orchestrated **ML workflow** with **AWS EC2** for model training, CVAT for data quality, and **AWS S3** data storage.
- Implemented **OpenCV video processing** for enhanced safety visualization, delivering **3x** quality assurance boost.
- Automated **ETL data pipeline** for **720+ hours** of video data from **Google Cloud to AWS S3** via **REST API**, significantly reducing operational time by **90%**.

Machine Learning Intern, *Sardar Patel Institute of Technology* **Feb 2021 – April 2021**
- Optimized **8 Machine Learning** models, namely Random Forest, Decision Trees, KNN, Logistic Regression, Support Vector Classifier, XGBoost, Lightgbm, & neural networks, resulting in **96%** accuracy.
- Trained **Stacked Ensemble ML model** for EEG emotion classification, achieving improved **97%** accuracy
- Conducted **Principal Component Analysis** to mitigate data's high dimensionality, reducing it by **94%**
- **Published research paper** in IEEE International Conference. DOI: 10.1109/ICCCNT51525.2021.9579818.

Data Scientist Intern, *Sardar Patel Institute of Technology* **Oct 2020 – Dec 2020**
- Led a team of five to engineer **3 Deep Learning CNN** models, namely ResNet50, VGG-16, and Inception V3 to aid Chest X-ray pneumonia detection with a maximum recall of **98%** and accuracy of **94%.**
- Employed **Image Processing** techniques, including **Data Augmentation**, to increase the data size by **5x.**
- **Authored a research paper** in IEEE International Conference. DOI: 10.1109/I2CT51068.2021.9417872.

Data Scientist/ Machine Learning Engineer, *Skinzy Software Solutions* **Oct 2020 – Jan 2021**
- Constructed a Mask-RCNN **Computer Vision** model in **TensorFlow** to detect skin diseases, achieving **0.6** IOU.
- Deployed a ResNet-50 **Deep Learning** model for disease **classification**, yielding **85%** accuracy by optimization.

Technical Skills

ML Libraries: TensorFlow, PyTorch, Pandas, Numpy, Scikit-Learn, Decision Trees, OpenCV, Unsupervised/ Supervised
Programming: Python, R, HTML, CSS, JavaScript, Flask | **Big Data**: PySpark, Databricks, Hadoop, ETL
Visualizations: Tableau, Power BI, D3.js, Matplotlib, Seaborn, Plotly, ggplot | **Tools**: AWS, Alteryx, Docker, Git, Linux
Databases: MySQL, MS SQL Server, Firebase, MongoDB, AWS S3, AWS RDS, AWS DynamoDB, PostgreSQL

Projects

Deep Learning models for Imbalanced Time Series Clinical data | *Deep Learning, Hyper tuning, TensorFlow*
- Experimented with **10 Deep Learning** methods to improve **time-series** performance in imbalanced data.
- Generated synthetic data with **SMOTE** & BorderLine SMOTE for overcoming **49%** data imbalance.
- Researched various Deep Learning methods, including **RNN** with echo state cell, **Transformer**, and **Feature Selection** using Random Forest Feature Ranking, resulting in a maximum AUC of **0.95%.**

Happiness Quotient Data Analysis | *Data Analysis, Management, Firebase, NoSQL, MySQL, Hadoop MapReduce*
- Managed **Firebase** & **SQL** distributed file storage for analyzing Happiness Index, GDP & unemployment.
- Engineered a web command-line interface in **Python and JavaScript** and **analyzed** user uploaded data for key trends like Top 10 GDP countries, Maximum Unemployment for Females in a given year.
- Deployed **Flask** website with **Hadoop-like MapReduce** for fast parallel data analysis & data search.

Big Data Yelp Business Recommendation System | *Data Mining, Big Data, Spark, JSON, Collaborative filtering*
- Built an **PySpark** Recommendation System for predicting rating for **1.5M** users and **200k** businesses.
- Executed **Item-based and ML-based** Collaborative filtering using **XGBoost Regression**, yielding **RMSE of 1.**
- Created an enhanced **hybrid recommendation system** with **data mining**, reducing **RMSE to 0.97**.