SHARVARI KALGUTKAR

Data Scientist

sharvari.kalgutkar82000@gmail.com (323) 470-3032 Los Angeles, CA Sharvari289 sharvarikalgutkar Portfolio

EDUCATION

Masters in Applied Data Science, *University of Southern California* **Aug 2022-May 2024**
Machine Learning, Data Mining, Data Management, Database Systems, Deep Learning, Data Visualization, Research Methods, Experimental Design and Analysis for User Studies, Fairness in AI (**CGPA 3.85/4**)

Bachelor of Technology Electronics and Telecommunication, *Sardar Patel Institute of Technology, India* **Aug 2018-Jun 2022**
Data Structures & Algorithms, Statistical Computational Lab, Object Oriented Programming, Applied Mathematics (**CGPA 9.52/10**)

TECHNICAL SKILLS

Machine Learning and Deep Learning: Python, R, TensorFlow, PyTorch, NumPy, Scikit-learn, SciPy, Matplotlib, Plotly, Seaborn, OpenCV
Tools and Technologies: PySpark, Databricks, Hadoop, Tableau, Power BI, D3.js, Amazon Web Services, Docker, CVAT, Streamlit
Databases and Infrastructure: SQL, Firebase, MongoDB, XML, Excel, AWS S3, AWS RDS, AWS DynamoDB, PostgreSQL, Linux, Git
Professional skills: Statistics, Data Mining, Unsupervised/Supervised ML, Data Visualization a nd Analysis, Computer Vision, Big data, NLP

PROFESSIONAL EXPERIENCE

Data Scientist, *CKIDS University of Southern California* **Feb 2024-Present**

- Research **neural network** forgetting, its impact on learning from **non-IID** data distributions and energy efficiency.
- Train and evaluate neural network models using **TensorFlow** in **distributed computing** environment such as **federated learning** versus **round-robin** training, assessing performance across both IID and non-IID datasets**.**

AI Engineer, *Scientist Technologies* **Nov 2021-May 2022**

- Cross-collaborated to develop five **Python**-based algorithms for road intersection safety analysis, achieving a **91% R2 score.**
- Implemented **OpenCV video processing** for enhanced safety visualization, delivering a **3x efficiency** boost in quality checks.
- Devised **Computer Vision** Road quality tracking system using models like **Faster R-CNN**, yielding highest **precision** of **84%.**
- Orchestrated an end-to-end **ML workflow**, leveraging **AWS EC2** for efficient model training, managing **data preprocessing** and annotation through **CVAT**, and storing data using **AWS S3**.
- Optimized **YOLOv5 object detection** model for diverse vehicle classification by integrating additional classes, resulting in **64% mAP**.
- Automated **data migration** of **720+ hours** from **Google Drive to AWS S3**, using **Google Cloud REST API**, drastically reducing time.

Deep Learning Engineer, *Skinzy Software Solutions* **Oct 2020-Jan 2021**

- Constructed a **Mask-RCNN instance segmentation** model in **TensorFlow** to detect skin diseases, achieving an **IOU of 0.6**.
- Deployed a **ResNet-50 Transfer** Learning model for skin disease classification, achieving **85% accuracy** through **optimization**.

PROJECTS

HappinessQ | *Data Analytics, Data Management, NoSQL, MySQL, Firebase, Hadoop MapReduce, Flask, JavaScript*

- Built **Firebase and SQL**-based **distributed file storage** for analyzing the World Happiness Index, GDP and unemployment.
- Engineered a web-based command-line interface in **Python** and **JavaScript** for manipulating user-uploaded files, enabling commands like directory creation, reading file partitions etc.
- Deployed a **Flask** website with **Hadoop**-like partition-based **MapReduce** for faster parallel data search and analysis.

Yelp Business Recommendation System | *Data Mining, Big Data, PySpark, Collaborative filtering*

- Built an **Apache Spark** Recommendation System for user-business rating prediction for **1.5M** users and **200k** businesses.
- Executed **Item-based** and **model-based Collaborative filtering** using **XGBoost**, yielding **RMSE** of **1.09** and **1**, respectively.
- Created an enhanced **hybrid recommendation system** with **feature mining**, reducing **RMSE** to **0.97**.

Deep Learning models for Imbalanced Time Series Clinical dataset | *Deep Learning, Model Hyper tuning, TensorFlow, Python*

- Experimented with **10 Deep Learning** methods to improve time-series performance in **highly imbalanced** medical dataset.
- Applied synthetic **data generation** with **SMOTE** and **BorderLine SMOTE** for improved imbalanced dataset handling.
- Researched varied techniques, including **SimplifiedRNN** and ModifiedRNN with echo state cell, **Transformer**, and advanced Ranking **Feature Selection** using **Random Forest Feature Ranking**, resulting in a maximum **AUC of 0.95%**.

Starbucks Stores Analysis | *Data Visualization, Dashboard, D3.js, Map box, HTML, CSS*

- Designed a dynamic **D3.js Dashboard** to analyze Starbucks' global store location strategy, KPI's optimizing decisions.
- Built a custom **Mapbox** Starbucks **store locator map** for LA, improving user navigation and accessibility to nearby stores.
- Executed global, country, and state-level analysis using **diverse data visualizations**, including Bar Charts, Scatterplots, Proportional Symbol maps, and Choropleth maps.

PUBLICATIONS

Pneumonia Detection from Chest X-ray using Transfer Learning (Team Lead*) |* *Deep Learning, Image processing, Data Augmentation*
EEG Brainwave Emotion Detection using Stacked Ensembling | *Neural Networks, ML, Predictive Analytics, SVM, Decision Tree*

AWARDS AND RECOGNITION

- Qualified as the **National Finalist** with a rank of **7 out of 600+** teams at the **Business Data Analytics** competition Anumaan, IIT Delhi in 2021 for analyzing key trends in sales data for six markets of an e-commerce company in a team of 3.