SHARVARI KALGUTKAR

Data Scientist, Machine Learning Engineer

skalgutk@usc.edu (323) 470-3032 Los Angeles, CA Sharvari289 sharvarikalgutkar Tableau

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 (**CGPA 3.8/4**)
Research and Methods and Analysis for User Studies

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

Python	Tools and Technologies	Databases	Professional skills
TensorFlow, Keras, PyTorch	Apache Spark	SQL, Firebase	Machine Learning, Statistics
OpenCV, Flask	Hadoop MapReduce	MongoDB, XML	Data Mining, Data Analysis
SciPy, Scikit-learn	Tableau, Power BI	AWS DynamoDB	Data Management, Data Visualization
Matplotlib, Plotly, Seaborn	AWS, Docker, CVAT	AWS RDS, AWS S3	Deep Learning, Computer

PROFESSIONAL EXPERIENCE

AI Engineer, Scientist AI Technologies Pvt Ltd**,** Internship **Nov 2021-May 2022**

- Engineered five Python Road intersection analysis algorithms (e.g., Time to Collision) in collaboration with transportation consultant team with an R2 score above 90%. Boosted safety visualization by integrating video processing using Open CV.
- Devised an automated road quality tracking system using Computer Vision models (Faster R-CNN, Cascade RCNN, RetinaNet, VF-Net). Utilized AWS EC2 for efficient model training. Managed data annotation through CVAT.
- Optimized vehicle classification with YOLO V5, adding more classes, achieving a 40% mAP score increase.
- Streamlined large data migration (Google Drive to AWS S3) via Google Drive API, empowering efficient data transfer.

Research Intern – Deep Learning and Machine Learning, Skinzy Software Solutions Pvt Ltd **Oct 2020-Jan 2021**

- Constructed a Mask-RCNN instance segmentation model in TensorFlow to detect skin diseases. Performed data annotation using the VGG annotation tool.
- Developed a histogram segmentation algorithm in Python to differentiate skin pixels from a background.

PROJECTS

DSCI 551: HappinessQ | Firebase, MySQL, Hadoop MapReduce, Flask, JavaScript.

- Built emulated distributed file storage (Firebase and MySQL) for global analysis (happiness, unemployment, GDP).
- Implemented EDFS commands (e.g., mkdir, readPartition) and stored file system metadata using Python and JavaScript.
- Deployed a Flask website for data search and analysis in EDFS, employing partition-based Hadoop MapReduce.

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

- Conducted an empirical study of 10 Deep Learning methods to improve imbalanced clinical time-series performance.
- Applied synthetic data generation with SMOTE and BorderLine SMOTE for improved imbalanced dataset handling.
- Analyzed varied techniques, including SimplifiedRNN and ModifiedRNN with echo state cell, Transformer, and advanced Ranking Feature Selection using Random Forest Feature Ranking, achieving a maximum AUC of 0.95% (test).

Yelp Restaurants Recommendation System |Data Mining, Spark RDD, Collaborative filtering

- Executed Item-based Collaborative filtering with Pearson Similarity and XGBRegressor recommendation using Spark RDD for user-business rating prediction, achieving an RMSE of 1.09 and 1, respectively.
- Created an enhanced hybrid recommendation system with feature mining, reducing RMSE to 0.9798 (test) & 0.9794 (val).

USC Campus Geospatial Data Analysis |Postgres, spatial databases, Google Earth Globe, Data visualization

- Visualized Campus attractions on Google Earth Globe by storing location coordinates in a KML file.
- Engineered a Postgres Database for spatial data storage, enabling computation of four nearest attractions and convex hull points from a starting location. Translated results into visualizations on Google Earth Globe.

PUBLICATIONS

EEG Brainwave Emotion Detection using Stacked Ensembling | DOI: 10.1109/ICCCNT51525.2021.9579818.

- Programmed a Stacked Ensemble model for emotion classification by combining outputs of deep neural networks and machine learning models, such as Random Forest, achieving 97% accuracy.

Pneumonia Detection from Chest X-ray using Transfer Learning |Team Leader **|** *DOI: 10.1109/I2CT51068.2021.9417872.*

- Led a team of five to engineer transfer learning CNN models (ResNet50, VGG-16, Inception V3) for chest X-ray pneumonia detection. Achieved 98.97% recall and 94.07% accuracy using image processing and augmentation.

ACTIVITIES

- Qualified as the National Finalist with a rank of 7 out of 600+ participants 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.
- Certifications: Udemy AWS Certified Machine Learning Specialty, DeepLearning.ai, Machine Learning University of Michigan.