Sharvari Kalgutkar

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

Education

Masters in Data Science, University of Southern California **Aug. 2022 – May 2024**
Data Analytics, Data Visualization, Database System, Machine Learning, Data Mining *(CGPA 3.85/4)*

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 – Present**

- Performed **visual data analysis** of 10+ data model's performance using **Python, Matplotlib, and Seaborn**, to enhance energy and cost efficiency, leading to **45%** cost improvement.
- **Presented stakeholders** with cost and energy comparisons of models using **line charts** and model **flowcharts**.

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

- 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%**.
- Cross-collaborated on **5 Python** algorithms for road safety **data analysis** using **Agile**, achieving a **91%** R2 score.
- Implemented **data visualization** using OpenCV video processing, delivering a **3x** quality assurance boost.
- Orchestrated ML workflow with **AWS EC2** for model training, CVAT for data quality, and **AWS S3** data storage.

Data Scientist/ Machine Learning Engineer, *Skinzy Software Solutions* **Oct 2020 – Jan 2021**

- Collected and **analyzed data** using **Python and Excel** for skin treatment recommendation system.
- Constructed a **CNN Computer Vision** model in **TensorFlow** to detect skin diseases, achieving an IOU of **0.6**.
- Deployed a ResNet-50 **Deep Learning** model for skin disease classification, yielding an accuracy of **85%**.

Technical Skills

Data Visualization: Tableau, Power BI, D3.js, Matplotlib, Seaborn, Plotly, ggplot, Excel
Programming: Python, R, HTML, CSS, JavaScript, Flask | **Big Data**: PySpark, Databricks, Hadoop, ETL
Data Libraries: Pandas, Numpy, Scikit-Learn, TensorFlow, PyTorch | **Tools**: AWS, GCP, Alteryx, Docker, Git, Linux
Datbases: MySQL, MS SQL Server, Firebase, MongoDB, AWS S3, AWS RDS, AWS DynamoDB, PostgreSQL

Projects

E-commerce Global Market Data Analysis | *Data Analysis, Data Visualization, Python, Matplotlib, Seaborn, Plotly*

- **National Finalist** with a rank of **7 out of 600+ teams** at the **Business Data Analytics** at IIT Delhi.
- Engineered **data visualizations** using Python, Matplotlib, Seaborn, and Plotly, featuring Barplots, Line Charts, Box Plots, Squarify plots, and World maps to analyze sales trends of 6 e-commerce markets.
- **Communicated to stakeholders**' seasonality trends, customer retention to identify top markets.

Starbucks Data Analysis | *Data Visualization, Statistical Analysis, Dashboard, D3.js, Mapbox, HTML, CSS*

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

Layoffs Dashboard | *Tableau*

- Link: https://public.tableau.com/app/profile/sharvari.kalgutkar/viz/NavigatingChangeTheLayoffsLandscape

Spotify Trending Hits Dashboard | *Tableau*

- Link: https://public.tableau.com/app/profile/sharvari.kalgutkar/viz/SpotifyTopCharts202

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**.