Vansh Rajesh Jain

213-814-7403 | Los Angeles, CA | vansh162000@gmail.com | vansh-jain16 | Vansh1610 | Portfolio

Education

Masters in Data Science, University of Southern California **Aug. 2022 – May 2024**
Database Systems, Machine Learning, Data Mining, Data Analysis and Visualization *CGPA 3.9/4*

Bachelors in Electronics Engineering, Sardar Patel Institute of Technology **Aug. 2018 – June 2022**
Data Structure, Algorithms, Object Oriented, Statistical Analysis, Management & Marketing *CGPA 3.9/4*

Experience (1+ years)

Data Scientist, *CKIDs University of Southern California* **Feb. 2024 – Present**

- Conducted **graphical analysis** on model performance with **Python, Matplotlib**, and **Seaborn** to visualize energy and cost efficiency, resulting in **45% cost reduction** among **10+** image data models.
- **Presented** comprehensive insights, model **flowcharts** and comparative analysis to a panel of **stakeholders.**

Computer Vision Engineer, *Dimensionless Technologies Pvt Ltd* **Dec. 2021 – May 2022**

- Conducted **data analysis** on penultimate layers to visualize model features, and enhanced the images using **OpenCV**, resulting in an **accuracy increment** of **1.2%.**
- Performed **ETL** to transform **JSON** to **CSV** from Google Drive to **Azure**, boosting **operational efficiency by 3x.**
- Implemented **data visualization** in **MS Excel**, incorporating ROC curve analysis, bar charts, and line charts.
- Executed **ML** algorithm using **agile** to optimize solar panel **false positives**, increasing precision by **24%.**

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

- Constructed a **CNN model** in **TensorFlow** to highlight skin abnormalities with an **IOU of 0.6.**
- Implemented a **deep learning** ResNet-50 model to detect skin abnormalities, yielding an **accuracy rate of 85%.**
- **Analyzed and organized** collected data using **Python** for skin disease treatment recommendation system.

Technical Skills

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

Projects

Starbucks Store Data Analysis Dashboard | *D3.js, JavaScript, HTML, CSS*

- Created a **D3.js dashboard** for **statistical analysis** of Starbucks store KPIs, aiding strategic decision-making.
- Developed dynamic **visualizations** using **JavaScript**, including Proportional Symbol Map, Dot Map, and Choropleth Map, to analyze store factors globally and for California.
- Integrated **Map Box** into the dashboard for enhanced visualization of store density in specific LA locations.

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

- Achieved **National Finalist** ranking among **600 teams** in the Anumaan Business Data Analytics competition.
- Analyzed sales across **6 global markets** with **Python and Matplotlib**, using Line charts, Squarify plots, and maps.
- Presented RFM analysis, Customer Retention, and Seasonal Patterns to stakeholders to identify the top market.

World Startups Data Analysis Dashboard | *Tableau*

- Tableau Link: https://public.tableau.com/app/profile/vansh.rajesh.jain/viz/GlobalStartupAnalysis/Dashboard1

Youtube Global Data Analysis Dashboard | *Tableau*

- Tableau Link: https://public.tableau.com/app/profile/vansh.rajesh.jain/viz/YoutubeGlobalAnalysis2023/Dashboard1

Happiness Quotient Data Analysis | *Data Management, MySQL, Firebase, MapReduce, Flask, NoSQL, Rest API*

- Created distributed storage with **MySQL** and **Firebase** for analyzing happiness, unemployment, and GDP.
- Built a **Flask** webapp for visualisation, alongside command-line tools using **Python** and **JavaScript** for retrieval.
- Employed **partition-based Hadoop MapReduce** techniques, including identifying top 10 GDP per capita countries and calculating mean freedom scores for efficient parallel analysis.

Yelp Review Big Data Recommendation System | *Spark, Machine Learning, XGBoost, Data Mining, JSON*

- Developed a **PySpark** recommendation system for Yelp, predicting ratings for **1.5M users and 200k businesses.**
- Built an **Item-Based Collaborative Filtering** and **XGBoost regression**, achieving an **RMSE of 1.09 and 1**.
- Constructed a **hybrid** recommendation model with **feature engineering**, resulting in an **RMSE of 0.97.**