VANSH RAJESH JAIN

Data Scientist

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

EDUCATION

Masters in Applied Data Science, *University of Southern California* (**CGPA 3.8/4**) **Aug 2022-May 2024**
Database Systems, Machine Learning for Data Science, Data Mining, Deep Learning, Los Angeles, USA
Data Visualization, Fairness in AI, Research Studies, Experimental Design, and Analysis for User Studies

Bachelors in Electronics and Telecommunication, *Sardar Patel Institute of Technology* (**CGPA 9.8/10)** **Aug 2018-Jun 2022**
2nd place silver medalist for outstanding performance during undergrad. Mumbai, India
Minor Introduction to Management and Marketing at S.P Jain Research and Management

TECHNICAL SKILLS

Machine Learning / Deep Learning: Python, R, TensorFlow, PyTorch, OpenCV, Scikit-Learn, Pandas, Matplotlib, Seaborn, NumPy, Plotly
Tools/Technologies: Pyspark, Hadoop, Databricks, Power BI, Tableau, Alteryx, Flask, Docker, D3.js, ETL, Amazon Web Services
Databases and Infrastructure: SQL, MongoDB, AWS RDS, AWS S3, AWS DynamoDB, Firebase, XML, PostgreSQL, Linux, Git, Excel
Professional Skills: Data Mining, Computer Vision, Data Analysis, Data Visualization, Big Data, NLP, Supervised/Unsupervised, Statistics

PROFESSIONAL EXPERIENCE (1 year)

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

- Research on how **neural networks** forget previous tasks and learnings on **non-IID data** while improving **energy efficiency.**
- Compare **distributed computing** methods such as **Federated and Round-Robin Learning** performance on Fashion MNIST **IID and non-IID** data distributions using **neural networks** in **TensorFlow**.

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

- Researched and showcased **seven object-tracking** algorithms in **OpenCV** for monitoring electronics on conveyors.
- **Collaborated with cross-functional** teams in an **agile environment** to train an **EfficientNet-B5 model** using **Docker** for distinguishing genuine from counterfeit electronics, achieving an impressive **accuracy rate of 97%.**
- Conducted **Grad-CAM analysis** on penultimate layers to visualize model features, resulting in an accuracy boost of **1.2%.**
- Trained separate **YOLOv4** models on **Azure GPU** to detect **15** solar panel defects, yielding an average **mAP of over 55%.**
- Implemented **penalty matrix** optimization in **Xgboost classifier,** performed **data cleaning** and enhanced image contrast with **OpenCV** to boost solar defect detection **precision** from **60% to 84%.**

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

- Constructed an **instance segmentation** model, **Mask RCNN** in **TensorFlow** to highlight skin abnormalities with an **IOU of 0.6**.
- Implemented a **Transfer Learning ResNet-50** model to detect skin abnormalities, yielding an **accuracy** rate of **85%.**

PROJECTS

HappinessQ | Data Management, Python, MySQL, Firebase, Hadoop MapReduce, Flask, JavaScript, NoSQL, Rest API

- Developed a **Flask** application to **analyze** global happiness using unemployment, GDP, and happiness index parameters.
- Created a file system using **MySQL and Firebase** to store CSV data in partitions for **distributed storage**.
- Executed command-line tools in **Python** and **JavaScript** for storage and retrieval and utilized **partition-based MapReduce** techniques for faster and more efficient parallel analysis.

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

- Conducted an in-depth **statistical analysis** of Starbucks to analyze store KPIs, facilitating strategic decision-making processes.
- Developed visualizations, such as **Proportional Symbol Map, Dot Map, and Choropleth Map**, using **D3.js and JavaScript** to illustrate Starbucks store distribution with population density, median income with number of stores, and competitor locations.
- Incorporated a **Map Box** into the dashboard to help users visualize the density of stores in specific locations.

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

- Achieved **National Finalist** ranking among **600 teams** in the Anumaan Business Data Analytics competition at IIT Delhi, **collaborating** with a **team of three** to analyze sales trends across six global e-commerce markets.
- Leveraged **Python and Matplotlib** to create visualizations such as **Line charts, Squarify plots, and Choropleth maps, incorporating RFM analysis, Customer Retention, and Seasonal Patterns** to determine the top-performing market.

Yelp Review Recommendation Systems | Python, Spark, Machine Learning, Xgboost, Data Mining, JSON

- Developed a **Spark**-based **Recommendation System** for Yelp data, predicting ratings for **1.5M users and 200k businesses**.
- Built an **Item-Based Collaborative Filtering** and **XGBoost model**, achieving an **RMSE of 1.09 and 1** on test data, respectively.
- Constructed a **hybrid** recommendation model with **feature engineering** techniques, resulting in an **RMSE of 0.97**.

PUBLICATIONS

EEG Brainwave Emotion Detection Using Stacked Ensembling | Predictive Modeling, Trees, Bagging, Boosting, Feature Selection

- **Led a team** to program a **stacked model** for emotion forecasting, achieving an **accuracy of 97%.**
- Combined 8 base models: **neural networks, Random Forest, SVM, Logistic Regression, KNN, XGBoost, and LightGBM**.
- Utilized **Principal Component Analysis** to **reduce dataset dimensionality by 94**%, optimizing computational resources.

Pneumonia Detection from Chest X-ray Using Transfer Learning |Image Processing, Transfer Learning, Data Augmentation

- Conducted a comparative study on **transfer learning models** including **VGG16, ResNet50, and Inception V3**.
- Trained models with **augmented images** to predict Pneumonia in Chest X-ray, achieving **recall of 98% and accuracy of 94%.**