Vansh Rajesh Jain

📞 213-814-7403 ◆ San Jose, CA ✉ vansh162000@gmail.com in vansh-jain16 ○ Vansh1610 ⊕ Portfolio

Education

Masters in Data Science, University of Southern California	**Aug. 2022 – May 2024**
Machine Learning, Data Mining, Deep Learning, Database Systems, Fairness in AI	*CGPA 3.87/4*
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 – May 2024**
- Researched **neural network** forgetting in distributed training like **FedAvg, Round Robin** on diverse data.
- Conducted **distributed model** performance study on **10+ deep learning** models to visualize their energy and cost efficiency, resulting in **45% cost reduction**.
- Recognized as the **'Best Data Science Team'** at **CKIDs USC** for exceptional **interdisciplinary research efforts.**

Computer Vision Engineer, *Dimensionless Technologies Pvt Ltd* — **Dec. 2021 – May 2022**
- Cross-collaborated to train EfficientNet-B5 **deep learning** model using **agile** on **Docker** distinguishing counterfeit electronics, achieving **97% accuracy.**
- Trained YOLOv4 **object detection models** for **15** solar panel defects on Azure, achieving **55% mAP.**
- Built a **machine learning** algorithm and **OpenCV** for image enhancement to optimize solar defect precision by **24%.**
- Conducted **Grad-CAM analysis** on final layers to visualize model features, boosting **accuracy by 1.2%.**
- Generated synthetic data using **GANs** for solar defect detection to decrease false positives.
- Performed **ETL** to transform **JSON** to **CSV** from Google Drive to **Azure**, boosting **operational efficiency by 3x.**

Machine Learning Engineer Intern, *Sardar Patel Institute of Technology* — **Feb. 2021 – Apr. 2021**
- **Led** a team to program stacked **predictive model** for EEG emotion detection, achieving **accuracy of 97%.**
- Combined **8 ML** models: neural networks, Random Forest, SVM, Logistic Regression, KNN, XGBoost, LightGBM.
- Utilized **Principal Component Analysis** to reduce dimensionality by **94%**, thus optimizing resources.
- **Published** research in the IEEE International Conference - **DOI: 10.1109/ICCCNT51525.2021.9579818 .**

Data Scientist / Machine Learning Engineer, *Skinzy Software Solutions Pvt Ltd* — **Oct. 2020 – Jan. 2021**
- Built **computer vision** Mask RCNN model in **TensorFlow** to highlight skin abnormalities with an **IOU of 0.6.**
- Implemented **deep learning** ResNet-50 model to detect skin abnormalities, yielding an **accuracy of 85%.**

Data Scientist Intern, *Sardar Patel Institute of Technology* — **Oct. 2020 – Dec. 2020**
- Conducted a study on **VGG16, ResNet50, and InceptionV3** for chest X-ray pneumonia prediction.
- Trained **deep learning CNNs** using **Transfer Learning**, achieving **98% recall** and **94% accuracy.**
- **Published** research in the IEEE International Conference - **DOI: 10.1109/I2CT51068.2021.9417872 .**

Technical Skills

Machine Learning: NumPy, Pandas, Scikit-learn, TensorFlow, Pytorch, Trees, OpenCV, Supervised/Unsupervised
Big Data: PySpark, Hadoop, Databricks, ETL | **Visualisations**: Matplotlib, Seaborn, Tableau, Power BI, D3.js, Plotly
Databases: MySQL, MS SQL Server, PostgreSQL, MongoDB, AWS RDS, AWS S3, DynamoDB, Firebase
Programming: Python, R, HTML, CSS, JavaScript | **Tools**: AWS, Azure, GCP, Alteryx, Flask, Docker, Excel, Git, Linux

Projects

Deep Learning for Imbalanced Time Series Clinical Data 🔗 | *TensorFlow, Deep Learning, Hyper tuning*
- Conducted a study to enhance imbalanced **Time Series classifiers** by integrating established methods.
- Developed **10 classification** models, incorporating Simplified **RNNs** with Echo State cells, **Transformers**, and Random Forest Feature Ranking, achieving a Test **AUC of 95%.**
- Integrated **SMOTE**, Borderline SMOTE, to tackle data imbalance, reducing data imbalance by **49%.**

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 visualization, 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.**