Vansh Rajesh Jain

213-814-7403 San Jose, CA vansh162000@gmail.com 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 computing** like FedAvg, Round Robin on diverse data.
- Conducted performance study on **10+ Deep Learning** models using **Tensorflow**, 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.**
- Utlilized **OpenCV** for image processing and built a **Machine Learning** model to optimize defect precision by **24%.**
- Conducted **Grad-CAM analysis** on **CNN** layers for model interpretability, boosting **accuracy by 1.2%.**
- Peformed synthetic data generation using **GANs** for solar defect detection, decreasing **false positives** by **8%**.
- Developed **ETL pipeline**, transforming **JSON** to **CSV** from Google Drive to **Azure**, boosting **efficiency by 3x.**

Machine Learning Engineer Intern, *Sardar Patel Institute of Technology* **Feb. 2021 – Apr. 2021**
- **Led** a team to program a **Stacked Ensemble ML Model** for EEG emotion detection by combining **8 ML** models: neural networks, Random Forest, SVM, Logistic Regression, KNN, XGBoost, LightGBM, achieving **97% accuracy.**
- 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 **deep learning CNNs** such as **VGG16, ResNet50, and InceptionV3** using **Transfer Learning** for chest X-ray pneumonia prediction, achieving **98% recall** and **94% accuracy.**
- Performed **Image processing** and **Data Augmentation**, to increase the data size by **5x.**
- **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

E-commerce Market Data Analysis | *Python, Matplotlib, Seaborn, Plotly, Data Analysis, Numpy, Pandas*
- Achieved **National Finalist** title, ranking **7th** among **600 teams** in IIT Business 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

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 for faster **parallel analysis**, including identifying top 10 GDP per capita countries and calculating mean freedom scores.

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.**