VANSH RAJESH JAIN

Machine Learning Engineer, Data Scientist

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

EDUCATION

Masters in Applied Data Science, *University of Southern California* (**CGPA 3.8/4**) **Aug 2022-May 2024**
Data Management, Machine Learning for Data Science, Data Mining, Deep Learning, Database Systems, Data Visualization, Fairness in AI — Los Angeles, USA

Bachelors in Electronics and Telecommunication, *Sardar Patel Institute of Technology* (**CGPA 9.8/10**) **Aug 2018-Jun 2022**
Data Structures and Algorithms, Object Oriented, Applied Mathematics, Statistics Computational Lab — Mumbai, India

TECHNICAL SKILLS

Machine Learning and Deep Learning: Python, TensorFlow, PyTorch, A/B testing, OpenCV, Scikit-Learn, SciPy, Matplotlib, Seaborn
Tools/Technologies: Hadoop, Databricks, Amazon Web Services, Power BI, Tableau, Flask, Apache Spark, Docker, D3.js, ETL, Linux
Databases: SQL, MongoDB, AWS RDS, AWS S3, AWS DynamoDB, Firebase, XML, PostgreSQL
Professional Skills: Data Management, Data Mining, Computer Vision, Data Analysis, Data Visualization, Big Data, NLP

PROFESSIONAL EXPERIENCE

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

- Research on how **artificial neural networks** forget previous tasks and learnings on **non-IID** data.
- Compare **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 to train an **EfficientNet-B5** model with **Docker on Linux** to differentiate genuine from counterfeit electronics, achieving **97% accuracy**.
- 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** and enhanced image contrast with **OpenCV** to boost solar defect detection **precision** from **60% to 84%.**
- Generated synthetic data for solar defect detection using a **GAN model** to further **decrease false positives**.

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 | Python, MySQL, Firebase, Hadoop MapReduce, Flask, JavaScript

- 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.

Deep Learning for Imbalanced Time Series Clinical Data | TensorFlow, Python, Jupyter Notebooks, Hypertuning, Numpy, Pandas

- Conducted a study to improve the performance of **imbalanced Time Series classifiers** by combining established methods.
- Built **10 classification models**, including **Simplified RNNs with Echo State cell** and **Transformers**, with a **Test AUC of 95%**.
- Integrated **SMOTE**, **Borderline SMOTE**, and **Random Forest Feature Ranking** to tackle data **imbalance** and model complexity.

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

- 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**.

USC Campus Geospatial Data Analysis | Spatial Database, Google Earth, Postgre, Data Visualization

- Performed **geospatial data analysis** of waterfalls and statues on the USC campus in **Google Earth**.
- Created a **KML file** to store coordinates and established a **PostgreSQL** database to store **spatial coordinates**.
- Executed spatial queries to visualize convex hull and identify four nearest neighbors from a starting point on Google Earth.

PUBLICATIONS

EEG Brainwave Emotion Detection Using Stacked Ensembling | *(Team Leader)* Deep Learning, Machine Learning, Python
Pneumonia Detection from Chest X-ray Using Transfer Learning | Python, TensorFlow, Transfer Learning, Data Augmentation

AWARDS AND RECOGNITION

- Secured a rank of **National Finalist** out of **600 teams** in the Business Data Analytics competition at IIT Delhi in 2021, in a team of 3, to **analyze sales trends** across six e-commerce markets.
- Acknowledged within the department as the **2nd place silver medalist** for outstanding performance during undergrad.