VANSH RAJESH JAIN

Machine Learning Engineer, Data Scientist

vanshraj@usc.edu (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 — Los Angeles, USA

B.Tech 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

Professional Skill	Tools and Technologies	Python	Database
Data Management, Data Mining	Hadoop Mapreduce	TensorFlow, Keras, PyTorch	MySQL, AWS RDS
Computer Vision, Deep Learning	PowerBI, Tableau	Flask	MongoDB, XML
Statistics, Machine Learning	AWS, Docker, D3.js	OpenCV, Scikit-learn, SciPy	Firebase
Data Analysis, Data Visualization	Apache Spark, Databricks	Seaborn, Matplotlib, Plotly	AWS DynamoDB, AWS S3

PROFESSIONAL EXPERIENCE

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

- Researched and presented 7 **object-tracking** algorithms in **OpenCV** for electronics monitoring on conveyors.
- Collaborated with **cross-functional team** and achieved **97% accuracy** in distinguishing genuine vs. counterfeit electronics using **GPU-accelerated** EfficientNet-B5 model within **Docker**, complemented by insightful Grad-CAM analysis.
- Created SolarAI software using **YoloV4** and **Azure's** GPU acceleration for solar defect detection and tracking of 15 defects.
- Improved defect detection from **60% to 84%** by combining penalty matrix, model scores, and Xgboost alongside enhanced image contrast using OpenCV techniques and utilized CVAT for **data annotation**.

Research Intern – Deep Learning and Analytics, *Skinzy Software Solutions Pvt Ltd* **Oct 2020-Jan 2021**

- Built an **instance segmentation** model for accurate skin abnormality detection and affected area highlighting.
- Applied Python **histogram segmentation** to distinguish skin color from other hues. Performed data annotation using VGG Annotator.

PROJECTS

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

- Built Emulated Distributed File System using MySQL and Firebase for global happiness, unemployment, and GDP analysis.
- Implemented EDFS command line tools in Python (such as rm, ls, put, getPartitionLoc, cat, etc.) for storing and retrieving data.
- Applied partition-based MapReduce techniques for data analysis and showcasing results through a Flask-based web application.

Deep Learning for Imbalanced Time Series Clinical Data | TensorFlow, Python, Deep Learning, Model Hypertuning

- Conducted an empirical study to elevate imbalanced clinical classifier performance by synergizing established methods.
- Employed varied techniques for time series classification, including Simplified RNNs with Echo State cell, Transformers, and advanced feature selection like Random Forest Feature Ranking for balanced data. (Total of 10 methods)
- Integrated SMOTE and Borderline SMOTE to address data imbalance, achieving a maximum test AUC of 95%.

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

- Implemented an Item-Based Collaborative Filtering with Pearson Similarity and XGBRegressor Model-based recommendation system in Spark on the Yelp dataset to predict ratings for users and businesses, achieving an RMSE of 1.09 and 1, respectively.
- Constructed a refined hybrid recommendation model combining above techniques by performing feature engineering, yielding RMSE of 0.9792 (validation) and 0.9798 (test).

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

- Generated a KML file to display waterfall and statue coordinates across campus on Google Earth.
- Developed a Postgre database to store spatial coordinates. Executed spatial queries to visualize convex hull and identify four nearest neighbors from a specified starting point on Google Earth.

PUBLICATIONS

EEG Brainwave Emotion Detection Using Stacked Ensembling | *(Team Leader)* | DOI: 10.1109/ICCCNT51525.2021.9579818

- Led a team to program a stacked model to forecast an emotion by combining outputs of eight base models (deep neural networks and machine learning models such as Random Forest) with a 97% accuracy.

Pneumonia Detection from Chest X-ray Using Transfer Learning | DOI: 10.1109/I2CT51068.2021.9417872

- Conducted a comparative study on VGG16, ResNet50, and Inception V3 transfer learning models trained with augmented images to predict Chest X-ray image Pneumonia, achieving a recall of 0.9897 and an accuracy of 0.9407.

ACTIVITIES

- Secured **7th position out of 582** participants 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.
- Completed Certification: Udemy AWS Certified Machine Learning Specialty 2023, DeepLearning.ai, Applied Machine Learning by the University of Michigan.