Sandeep Kamal Jalui

Gainesville, Florida • sandeepjalui5@gmail.com • 352-283-4866 • linkedin.com/in/sandeep-jalui-8a9290131/ • github.com/Sandeepjalui93

EDUCATION

University of Florida, Gainesville, FL — Aug 2021 - May 2023
MS in Electrical and Computer Engineering (Thesis in ML/CV) — GPA: 3.36/4.00
Relevant coursework: Fundamentals of Machine Learning, Applied Machine Learning, Deep Learning for Computer Graphics

University of Mumbai, Mumbai, India — June 2014 - June 2018
Bachelor of Engineering in Electronics and Telecommunication Engineering — GPA: 7.75/10

PROFESSIONAL EXPERIENCE

University of Florida - Gainesville, FL — Jan 2022 - May 2023
Graduate Research Assistant – Deep Learning Researcher
• Pioneered research methodologies on state-of-the-art architecture to do Instance Segmentation on 3D point clouds data specific to Oil and Gas industry which classified eight objects in an entire scenario.
• Initiated with my professor's research work, a neural network-based model was first implemented on the UF Hypercomputing system using 3D CNN algorithms and CUDA, which achieved a fantastic 90% accuracy on the test dataset, the first benchmark achieved.
• Conceptualized and innovated a new algorithm pipeline based on Graphical Convolution Neural Network by considering Point clouds as a Graphical Network of data for which the training speed was increased by 2x than the previous neural network model and accuracy in the range of 80% - 85%.

Tadaa.ai - Santa Clara, CA — Jan 2021 - July 2021
Machine Learning Engineer
• Researched and implemented several time series and regression-based models on finance datasets to forecast and predict financial revenue data for clothing company GAP as compared to earlier traditional financial techniques, machine learning models helped in giving faster and more accurate results for revenue forecasting.
• Higher accuracy was achieved on boosting models like LightGBM which gave 80% accuracy on the test dataset as compared to time series models which was below 60%.
• Lead the project and built AutoML forecasting software system which will take Excel data as input and give prediction and analytical insights from the data by running LightGBM algorithm in the backend. The average accuracy was in the range of 70% - 80% based on the different types of input data fed into the system.

Markytics - Mumbai, India — Oct 2020 - Dec 2020
Data Science Intern
• Supervised closely with the client company Aditya Birla Finance Limited to assist in bug fixing and feature improvement in the backend software which doubled the performance of their software using the Django framework and Rest API tools.
• Built Python scripts for scrapping pdf files, data pre-processing, and data extraction to obtain a clean dataset and visualized key trends using Statistical Frameworks and Power Bi which helped the business to understand their customer's loan approval process.
• Developed a Full-stack website using the Django framework, which helped provide information about a company's stock portfolio and analytical insights using the Yahoo finance library.

PROJECTS

Credit Card Fraud Detection System using Classification Models. — April 2023 - May 2023
• Data analysis using statistical tools and data preprocessing using Pandas and NumPy library were performed on the highly unbalanced dataset with 284,807 records to make the data ready for modeling.
• Several Classification model pipelines were built and tested using sklearn library to predict credit fraudulent cases with the model accuracy range of 91% - 99%.
• Logistic, Linear Discriminant Analysis and Gaussian NB gave an accuracy of about 91%, while models like Decision Tree and Random Forest overfitted the data by giving an accuracy of 99% hence I resolved the issue by using cross-validation techniques to reduce the overfitting of the models, thus giving accuracy around 95%.

Supermarket Sales Data Analytics using Regression and Classification Models. — Sept 2022 - Oct 2022
• Data analytics and statistics visualization plots were developed on the supermarket datasets showing the comparative performance of 3 markets using Matplotlib and Seaborn. These analytics gave insights into the product sales overview in 3 different markets in 3 different cities.
• Regression model pipeline was performed to predict the unit price and gross income of 3 supermarkets with an average accuracy of more than 95% and plotted stem figures to identify the parameters impacting the output.
• Classification model pipeline was formulated to predict the gender & customer type with average accuracy above 90% using cross-validation and grid search techniques.

Face Colourization using Auto Encoders and Transfer Learning. — Sept 2022 - Oct 2022
• Regressor model was developed using Neural Networks to predict the mean chrominance channels a and b using L channel as input with MSE of the order 10^{-5}.
• Based on the regressor model, the colorizer model was developed on the B/W face image dataset and then transferred to the fruits and vegetable images dataset using Auto Encoders and the Transfer Learning concept giving the average model accuracy of 85%.
• GPU CUDA was integrated with this project which was 1.7x faster than the CPU.

Car position detection with bounding box using Convolutional Neural Network with TensorFlow and Transfer Learning. — Nov 2022 - Dec 2022
• Written a CNN code for car object detection in the image using the TensorFlow library with average model accuracy in the range of 80% - 82%.
• Built a Neural Network algorithm on the car image dataset to identify the car positions in the images using Xception as the base pre-trained model and further neural network layers were added to build the model to predict the 4 coordinates of the bounding box around the car object.
• mIoU score was used to evaluate the model with the highest train score of 85.29%, validation score of 78.35%, and test score of 67.66%.

Handwritten Symbol Classification using Convolutional Neural Network with PyTorch. — Oct 2021 - Dec 2021
• Developed a CNN algorithm using PyTorch library on the handwritten symbol dataset consisting of 23,000 handwritten symbol images containing 25 classes/symbols.
• The training model achieved 98% accuracy and 96% accuracy on the blind/test dataset.
• GPU CUDA code was implemented along with the neural network code to improve the performance by 1.4x times than CPU.

TECHNICAL SKILLS

Programming Languages: Python, C, C++, CUDA, MySQL, PostgreSQL, JavaScript, CSS3, HTML5.
Machine Learning: Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy, SciPy, NLTK, XgBoost, LightGBM, OpenCV, Point Cloud.
Deep Learning: ANN, CNN, RNN, LSTM, GNN, GAN, Auto-Encoders.
Time Series models: AR, MA, ARMA, ARIMA, SARIMA.
Data Visualization: PowerBI, Tableau, Matplotlib, Seaborn, ggplot2, MS Excel.
Other Skills: JIRA, AWS, Linux, Git, Docker, Kubernetes, Django, Flask.

PUBLICATIONS AND ACTIVITIES

Attended "The 16th Learning and Intelligent Optimization Conference" in Milos, Greece. — June 2022
• Gave a presentation with the topic "Investigation of Graph Neural Networks for Instance Segmentation of Industrial Point Cloud Data" at the conference published by Springer-Verlag in LNCS.
Attended "2022 European Conference on Computing in Construction" in Rhodes, Greece. — July 2022
• Gave a small talk on my lab's project with the topic "Synthetic Point Cloud Generation for Class Segmentation Applications" at the conference published by EC3.
Hackathon participation hosted by Nvidia and UF Hypergator. — April 2022
• Model development and training of 3D Point clouds classification using Graphical Neural Network. Training speed was achieved 1.5x faster using Nvidia singularity docker.
Took a few classes as a Teaching Assistant teaching basic Machine learning algorithms for the construction school. — Oct 2022