Nikunj Lad

nikunjrlad@gmail.com | 857.999.6620 | **https://nikunjlad.dev**

SUMMARY

Data Scientist with 2.5 yrs of hands-on experience using Python, Pandas and NumPy while handling large-scale structured & unstructured datasets using Scikit-Learn, XGBoost and Keras. I have experience performing statistical analysis on data right from hypothesis testing, data preprocessing and debiasing to multi-core model training

EDUCATION

NORTHEASTERN UNIVERSITY   Sep 2018 - Aug 2020
MS in Information Systems, Concentration: Machine Learning & Pattern Recognition, **GPA: 3.8/4.0**
Coursework: Machine Learning, Big-Data Intelligence & Analytics, Data Structures & Algorithms, Parallel Computing

K. J. SOMAIYA COLLEGE OF ENGINEERING   Jun 2012 - May 2016
BE in Electronics and Telecommunications Engineering, Concentration: Image Processing

SKILLS

Languages: Python, MATLAB, C++, BASH, Markdown
Frameworks & APIs: Pandas, NumPy, Matplotlib, Plotly, Scikit-Learn, XGBoost, CatBoost, Keras, PyTorch, Tensorflow, Discovery HPC, NLTK, SQL, MongoDB, AutoML, H2O, AIF360, OpenCV, SciPy, PySpark, MapReduce
SDLC | VCS | Cloud: Agile, Waterfall, Iterative, Git (Bitbucket, GitHub), CircleCI, Docker, Amazon AWS, Terraform
Algorithms: Supervised & Unsupervised Learning, Outlier Detection, Dimensionality Reduction, Bias-Fairness Detection, Feature Selection, SMOTE Analysis, Hypothesis Testing, Hyperparameter Optimization

EXPERIENCE

UII AMERICA INC. | Computer Vision Research Intern   Sep 2019 – Jan 2020
- Implemented Dense Correspondence Estimation algorithm for real-time patient inference and increased inference speed by **50%** to **21 FPS** using multi-threaded Dockerized architecture on Nvidia P5000 GPU
- Worked on skeletal tracking in human body by utilizing Astra Pro Depth camera and annotated RGBD images in C++ for obtaining end-to-end mesh recovery of human body on NVidia RTX 2080 GPU

NORTHEASTERN UNIVERSITY | Graduate Teaching & Research Assistant   Jan 2019 – Apr 2020
- Conducted Deep Learning workshops for a batch of **100 - 150** students on Convolutional Neural Networks
- Reduced run-time complexity by **60%** by developing micro-services architecture using H2O automodeling framework
- Memory and time profiled metrics on Tensorboard using PyTorch and achieved **2x** faster dataloading with Dask

INVIDEO | Software Engineer   Sep 2017 – May 2018
- Scaled up video production rate for clients from **30 videos/day** to **300 videos/day** by implementing Geometric Perspective Transformation on image masks
- Improved software efficiency by **60%** by saving **10 mins** of rendering time per video by incorporating Face Detection using Convolutional Neural Networks for automated text positioning
- Parsed new articles for understanding language semantics using Spacy and NLTK for text-to-video translation

PROJECTS

VISUALIZING BIAS-FAIRNESS IN AI USING IBM AIF360 *(Python, AIF360)*
- Currently observing and visualizing Bias and ensuring Fairness in machine learning systems using Statistical Parity Difference, Equal Opportunity Difference, Average Odds Difference, Disparate Impact and Theil Index Metrics
- Mitigating Bias and ensuring fairness using optimized pre-processing & adversarial debiasing algorithms
- Also working on reducing data dimension using t-Stochastic Neighbor Embedding (t-SNE) and developing interpretable models using Local Interpretable Model Agnostic Explanations (LIME)

AVAZU AD CLICK-THROUGH RATE(CTR) PREDICTION *(XGBoost, Scikit-Learn, Plotly)*
- Performed statistical tests for feature significance using P-Values and Variance Inflation Factor (VIF) analysis
- Removed outliers using Isolation Forest and Minimum Covariance Determinant (MCD) followed by multi-processing based SMOTE Analysis for solving Class Imbalance problem
- Applied Logistic Regression, XGBoost, CatBoost and Multi-Layer Perceptron(MLP) for modeling the processed data and achieved an average accuracy of 81% on out of sample data with a 95% Confidence Interval of 0.00545

KING COUNTY HOUSING PRICE PREDICTION *(XGBoost, Scikit-Learn, Python, Pandas)*
- Performed Normality tests using QQ-Plots and maintained homocedastic nature of the variables
- Automatically Selected best features using K-Best fit using ANOVA and Variance Thresholding
- Implemented Generalized Linear Models, Gradient Boosting Machines and Random Forests for housing price prediction with Lasso and Ridge Regression and evaluated models using R-Squared and Adjusted R-Squared metrics

PUBLICATIONS

Nikunj R. Lad, J. H. Nirmal, Kshipra D. Naikare. **Total variability factor analysis for dysphonia detection**. *International Journal of Information Technology - Springer*. 2019, Vol. 11, Issue 1, pp 67–74 *(Won Best Paper Award)*