Prerna Yadav

New York, NY 10038 | prernay1314@gmail.com | (551)-260-4890 | GitHub | LinkedIn | prernayadav.github.io

Professional and Academic Experience

Machine Learning Intern | *Seidenberg School of CS/IS, Pace University* Sep 2023 - Present

- Built a high dimensional (400K+, 100+) classification **Neural Net** model in **R** on child welfare data for child risk identification.
- Conducted feature selection & engineering via **SHAP**, permutation importance & **K-Means** clustering of related variables.
- Applied locally estimated scatterplot smoothing (LOESS) and Isotonic calibration for a better Precision-Recall ratio at 10 different prevalence thresholds that better account for child risk identification probabilities.
- Introduced the PR-Curve as a classification metric to accurately measure performance from imbalanced data.

Data Analyst | *Boss Program, OASIS, Pace University* Mar 2023 - Aug 2023

- Researched factual association characteristics on knowledge storage within **GPT**-models through GPT-2XL.
- Traced relations and knowledge tuples to detect specific neuron activations within **MLP** mid-layers with **ROME** method.

Information Technology Services Intern | *EdMedia, Pace University* Aug 2023 - Present

- Assisted tier 1-3 support for 500+ users across **Windows, Linux,** and **Android** devices. Diagnosed complex software issues.
- Configured & installed over 200 software apps, operating systems suites (**Microsoft Office**, **G Suite**) on devices.
- Caught over five thousand defects and bugs which helped to decrease the number of bugs found in production by 25%.

Data Analyst | *Infotech* Mar 2020 - May 2022

- Extracted, interpreted, and analyzed data to identify key metrics and transform raw data into meaningful, actionable information.
- Performed data analysis using **Excel**, providing routine operational reports/dashboards with metrics for management.
- Demonstrated utilizing **ETL** tools using PL/**SQL** server integration services (**SSIS**). Automated data fetching using **Python.** scripts.

Data Applications

Data Analyst | NYC Crime Analysis | *R, ML- Random Forest Model* Analysis | GitHub

- Deployed **R** for encompassing Data Collection, Data Preprocessing, Statistical Analysis, Predictive modeling, and Visualization.
- Analyzed, **dplyr, ggplot2, Random Forest**, and **leaflet** to extract insights, identify crime trends, and create interactive visualizations, offering valuable tools for law enforcement and community safety, with accuracy of 74.8%.

Data Scientist | Automatic Bearing Defect Inspection System | *Python, ML, Computer Vision,* | [*SIH - (SKF)]*

- Achieved high-precision defect detection through real-time defect classification with Machine Learning algorithms.
- Conceptualized Computer vision, **Python** & **tkinter** with an efficient data centralization mechanism and enriched with visual analytics via **Matplotlib** ensuring remarkable accuracy with 95.6% accuracy.

Data Analyst | Titanic Survival Prediction |*Python, ML, Dataset - Kaggle* Analysis | GitHub

- Prepared a **Logistic Regression** model to analyze and predict the survival of the Titanic passengers from the Kaggle.
- Generated exploratory data analysis, and **mean imputation** using Pandas and NumPy. Execute one-hot encoding, K-fold cross-validation, and Hyperparameter Tuning, improving the model's accuracy from 62% to 77%.

Education

Master of Science (MS) in Information Systems | Concentration: Data Science **May 2024**
Pace University, Seidenberg School of CS and IS | New York.
Courses: NLP, Forecasting/Time Series, Deep/Machine Learning, Probability Theory, Statistical Inference, Algorithms, EDA

Bachelors in Computer Science | University of Mumbai **Feb 2021**

Technical Skills and Tools

Tools/Frameworks: Python, R, SQL, Tableau, PowerBI, GCP BigQuery, AWS, AA (RPA), PySpark, PyTorch
Algorithms: Prophet, DeepAR, SARIMA, NN, Boosting, RandomForest, SVMs, kNN, Regression, L1/L2 Regularization, PCA
Packages*:* torch, scikit-learn, pandas, NumPy, sweetviz, seaborn, statsmodels, tidyverse, h2o.ai, ggplot2, dplyr

Publications

Paper Published: Vol. 5, Issue 10, October 2016 – *ISSN: 2278-1021* - TEES: Search Scheme over Encrypted Data
Volume: 04 Issue: 03 | Mar-2017 – *ISSN: 2395 -0056* - Web of Short URL's