# VIKAS RAWAT

Flushing, NY 11354 | vikas.rawat@pace.edu| 646-875-7176 | LinkedIn | GitHub

**ASPIRING DATA SCIENTIST | DATA ANALYTICS & VISUALIZATION | AI / ML ENTHUSIAST | SKILLED IN PYTHON SCRIPTING**

## EDUCATION

**Pace University, Seidenberg School of Computer Science and Information Systems** — New York, NY
Master of Science (MS) in Data Science | **GPA:** 3.96 — May 2025

**Indira Gandhi National Open University** — Delhi, India
Bachelor of Arts (BA) in Psychology — June 2023

## TECHNICAL SKILLS AND CERTIFICATIONS

**Programming Languages:** Python, R, C, C++, SQL, Bash | **Database Management:** SQL, MySQL, SQLite, MongoDB, HBase
**Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, fastai, PyTorch, NLTK, FastAPI, Flask
**Machine Learning Algorithms**: Linear Regression, Logistic Regression, Decision Tree, Random Forest, KNN, K-means, PCA, Neural Networks, CNN, Collaborative Filtering, RNN, Natural Language Processing (NLP), Large Language Models (LLM), Retrieval Augmented Generation (RAG)
**Miscellaneous:** AWS, Hadoop MapReduce, Git, GitHub, MS Excel
**Virtual Internship:** Accenture, Data Analytics and Visualization Job Simulation, The Forage
**Certifications**: Supervised Machine Learning (Coursera), CS50P: Python (Harvard), CS50x: Intro to Computer Science (Harvard)
**Hackathons:** Computer Vision Hackathon, Voxel51 (Winner) | The Merge Hackathon, Causality (2nd Place)

## EXPERIENCE

**College and Beyond Inc** — Remote
Data Intern — June 2024 – July 2024

- Lead rigorous data cleaning and preprocessing of college and student data using Pandas, from 5+ sources, to ensure accuracy and consistency across college rankings, admission statistics, and student performance metrics
- Crafted compelling data visualizations using Matplotlib and Seaborn, including Box plots, Heat maps and histograms, revealing insights on college location trends, tuition fees, and academic performance
- Collaborated in team meetings and brainstorming sessions, to propose data-driven strategies for enhancing student engagement and retention

## PROJECTS

**Fraud Alert** — December, 2024

- Developed a Flask-based system leveraging AssemblyAI to transcribe phone calls, detect sensitive keywords (e.g., bank details), and alert users to prevent potential scams.

**Recommendation System using Nearest Neighbors Machine Learning Algorithm** — May, 2024

- Performed data cleaning and EDA, merging and aggregating DataFrames in pandas, removing inconsistencies and duplicates
- Implemented a utility matrix for collaborative filtering, resulting in a movie recommender system

**Anomaly Detection using EllipticEnvelope and IsolationForest Algorithm** — March, 2024

- Visualized timeseries data of an instrument using plotly and Dash, and performed datetime feature extraction, enabling comprehensive anomaly detection
- Implemented dimensionality reduction, increasing the silhouette score by 92.6%, enabling robust clustering
- Performed anomaly detection on the reduced data using 2 different algorithms and 2 different techniques

**Stock Price Prediction** — February, 2024

- Collected and analyzed Apple stock price data using the yfinance API, totaling over 8 years of data
- Engineered features, including a binary trading signal from daily price movements and a comparison of 50-day vs 200-day moving averages, using NumPy and pandas
- Evaluated multiple classification algorithms, including K-Neighbors, Random Forest and XGBoost, to predict price direction
- Implemented a simpler model based on moving averages comparison without machine learning, improving accuracy by 9%, illustrating the importance of strategic feature selection