VED SHUKLA

vshukl01@nyit.edu | https://github.com/vshukl01 | (551)-655-5647 | https://vshukl01.github.io/portfolio/| **www.linkedin.com/in/ved-shukla-vs1710** | New York City

EDUCATION

New York Institute of Technology *M.S. in Data Science (GPA – 3.8)* *Sept 2023 – May 2025*

Relevant Coursework: Data Structure and Algorithm, Optimization Methods for Data Science, Machine Learning, Deep Learning, Data Base Management Systems, Artificial Intelligence, Statistical Concepts for Data Science.

SKILLS

Programming Languages: Python (Numpy, Pandas, Matplotlib, Scikit-Learn, NLTK, Ggplot, Plotly, PySpark, Seaborn, Vega-Altair, Beautiful Soup, Selenium), SQL (DDL & DML), C, C++, R, HTML, MATLAB

IDE: Anaconda, GIT, PyCharm, VS Code, Google Collab

ML & AI: Machine Learning (Linear Regression, Logistic Regression, K-means Clustering, PCA, XGBoost), Deep Learning (TensorFlow, PyTorch), Natural Language Processing (Transformers, GPT), Large Language Models (LLM), Time Series

Data Analytics and Visualization: Data Cleaning, Data Transformation, Data Modeling, Statistical Analysis, Data Visualization (Tableau, PowerBI)

Data Engineering & Database: Cloud **(**AWS, Google Cloud, Snowflake**),** MySQL, PostgreSQL, ETI. Pipelines, Data Integration, Docker

MS Skills: Excel, Word, PowerPoint

Organizational Skills: Perseverance, Attention to detail, Communication, Multitasking, Problem Solving, Teamwork

WORK EXPERIENCE

Data Analyst [Academic Technology Services – New York Tech] *June 2024 – Present*

- Developed and managed ETL pipelines using **Apache Airflow**, **PostgreSQL, and Python** (Pandas) to integrate and clean diverse datasets extracted from the backend with **L**earning **M**anagement **S**ervices platform-reported metrics APIs in real-time, ensuring accurate attendance tracking, survey analysis, participation, and performance level.
- Designed real-time **Tableau dashboards** for instructor, student, and Department level and performance monitoring, automating data reporting workflows with Data Build Tool, reducing manual data transformation efforts by 40%, and enabling data-driven strategic planning.
- Utilized **GitHub** for version control, improving collaboration and safeguarding the learning management system's codebase, ensuring seamless project development.
- Created **Functional Requirements Documentation (FRD)** to align system requirements with organizational system availability, ensuring a smooth migration and implementation process.

Graduate Assistant [Academic Success and Enrichment – New York Tech] *January 2024 – Present*

- Analyzed tutoring data using **Slate queries and Tableau**, providing key insights that optimized student support and resource allocation.
- **Led hiring** and scheduling for the **Supplemental Instructor Project**, improving lesson plan tracking, attendance management, and student-tutor engagement efficiency.
- Redesigned the **departmental website**, improving user experience, navigation, and accessibility, making academic support services more accessible.
- Developed process documentation and instructional videos, ensuring continuity and streamlined onboarding for future staff.

Graduate Assistant [Career Success and Experiential Education – New York Tech] *January 2024 – May 2024*

- **Streamlined paperwork procedures** by implementing standardized protocols and personally guiding students through the process, effectively **alleviating administrative processes** while enhancing service delivery.
- **Tailored career advising**, enhancing student profiles for a professional presence. **Conducted workshops** and presentations on resumes, cover letter writing, career guidance, and LinkedIn optimization.

PROJECTS

Intelligent Water Quality Monitoring:

- Processed **360GB** of test data, leveraging **ML models (Random Forest, ANN)** to predict 24-hour turbidity levels with **>95%** accuracy, significantly reducing dependency on manual water testing.
- Developed a model to **predict water contamination levels** with changing seasonal patterns, using **Agglomerative Clustering with PCA**, significantly reducing the manual data review time and **improving anomaly detection** accuracy by **35%**, ensuring proactive water quality management.

Data-Driven Mobile User Marketing Campaign Optimization:

- Identified OS-based consumer trends from **1TB** of user behavior data, showing iOS users install 15% more apps but use 20% less daily data than Android users, enabling **personalized app recommendations** and **tailored marketing strategies**.
- Automated user segmentation using Machine learning algorithms: **K-Means Clustering, Decision Trees, and Apriori Association Rules**, allowing for faster detection of emerging user trends, enabling **real-time adaptation of marketing strategies** to **optimize ad targeting** and engagement, improving efficiency by **30%**.

NYC Rideshare Analysis Using PySpark:

- Processed **50GB** of NYC rideshare data using **PySpark, applying Gradient Boosted Regression** to achieve an R-squared of **0.7** and uncover key demand trends.
- Implemented the **PageRank algorithm** in **PySpark** to analyze peak activity regions by time of day, providing actionable insights into rider demand patterns.

BorderMetrics: Predictive Border Intelligence and Visualization

- Engineered a predictive analytics pipeline using Python, SQL, Tableau, and Scikit-learn to process and model 20GB+ of U.S. border entry data spanning air, road, and water transport modes, enabling granular insight into seasonal and modal traffic behaviors.
- Delivered 10-year traffic forecasts, anomaly detection, and high-risk port classification, driving data-informed optimization of national border infrastructure, customs operations, and cross-border mobility planning.

LEADERSHIP & EXTRA-CURRICULAR

- **Orientation Leader**: Led orientation sessions for new students, guiding them through academics and campus resources while fostering a welcoming environment.
- As **President & Founder of the Data Science Club**, organized events and workshops, increasing membership by **40%** and promoting hands-on learning.