

SHASHANK SHEKHAR

New Jersey, NJ | +1(703)-930-2509 | sshekha1@mail.yu.edu | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

TECHNICAL SKILLS

- **Languages:** Python, SQL, C#
- **Machine Learning:** scikit-learn, TensorFlow, Statistical Modeling, Unsupervised Learning, KNN, SVM, Random Forest, Decision Tree, Naive Bayes, NLP
- **Data Analysis:** NumPy, Pandas, EDA, Feature Engineering,
- **Visualization:** Matplotlib, Seaborn, Tableau, Storytelling
- **Databases & Tools :** MSSQL, PostgreSQL, Oracle , Git, REST API, AWS

WORK EXPERIENCE

Harman Connected Services Pvt. Ltd.

Apr 2022 - Feb 2023

Associate Software Engineer

Pune, India

- Engineered and customized product **ImedOne's** features using C# and .NET, boosting client satisfaction by 10% through improved workflows.
- Executed predictive analysis on 50,000 mental health records using Python and scikit-learn, SVM model achieved 91% precision.
- Enhanced diagnostic accuracy for high-risk groups, cutting null error rate from 69.34% to 24%.

PROJECTS

Disease Data Modeling and Warehousing Project | PostgreSQL, ETL, Data Modeling, AWS Architecture, DbSchema

- Architected a scalable data model in PostgreSQL with a dimensional structure and ETL pipelines to optimize disease data storage and streamline data flow.
- Created ER diagrams using DbSchema and built dashboards for visualizing disease trends, enhancing healthcare decision-making.
- Proposed AWS-based architecture, comparing Snowflake vs. PostgreSQL for performance and cost.

Dropout Rate Prediction | Python, k-Fold Cross-Validation, Confusion Matrix, Random Forest, Decision Tree, Statistical Modeling

- Constructed the Random Forest model with 87% AUC, outperforming Decision Tree's 79% AUC, with 69% precision and 68% recall on imbalanced classes.
- Tested two feature sets, with the 23-attribute Random Forest model achieving 87% AUC and 84% accuracy, surpassing Decision Tree performance.
- Analyzed 73,000 records to uncover trends like higher dropout rates in larger schools and more Regents diplomas in low-needs districts, providing actionable insights.

E-commerce Purchase Prediction | Clustering, KNN Modeling, Python, Data Segmentation, Predictive Analytics

- Built a KNN model with 97.13% accuracy on e-commerce data, identifying three customer segments
- Identified three customer segments (high-value, moderate, low-engagement) with 12 optimized features to enable data-driven, targeted marketing strategies.
- Enabled targeted marketing strategies to boost conversions through data-driven segmentation.

Heart Disease Prediction Modelling | Python, SVM, Random Forest

- Examined heart disease data of 1.3M patients with high probability of experience heart attacks in future.
- Predicted patients at highest risk by developing machine learning models (Random Forest, SVM) with 92 % accuracy.

NYPD Calls for Service Data | Tableau, Storytelling

- Researched NYPD calls for service data in Tableau, revealing Brooklyn as the highest-crime borough with peak incidents from 12 am to 3 am.
- Delivered findings to show borough-specific crime hotspots and time-based patterns, enabling targeted decision-making and resource planning.

EDUCATION

Katz School of Science & Health - Yeshiva University

Jan 2024 - Expected May 2025 | New York, NY

Master of Science in Data Analytics & Visualization

Cumulative GPA: 3.5

Relevant Coursework: Machine Learning, Data Science, Structured Data Management, Computational Math's & Statistics, Storytelling

Guru Nanak Institute of Technology

May 2018 - May 2022 | Kolkata, INDIA

Bachelor of Technology in Computer Science Engineering

GPA: 3.6

CERTIFICATIONS

IIT Bangalore

Apr 2023 - Dec 2023 | Bangalore,INDIA

Advance Certificate Programme in Data Science

GPA: 3.4