

JAYASHREE JOHNSON (She/Her)

One Pace Plaza, NY, 10038 | jayashree.j@pace.edu | 201-736-4914 | [LinkedIn](#) | [GitHub](#)

PROFILE

Graduate student pursuing a Master of Science (MS) in Data Science. Proficient in Python, SQL, and Machine Learning. Experienced in exploratory data analysis, ML algorithms, DBMS, & predictive modeling, and building end-to-end data pipelines for real-world applications.

EDUCATION

| | |
|---|----------------------------|
| Pace University, Seidenberg School of Computer Science and Information Systems | New York, NY |
| Master of Science (MS) in Data Science GPA: 3.96/4 | May 2026 |
| Justice Basheer Ahmed Sayeed College for Women | Chennai, Tamil Nadu, India |
| Bachelor of Science (BS) in Advanced Zoology & Biotechnology GPA: 8.1/10 | May 2024 |

RELEVANT COURSEWORK

Python Programming | Mathematical Foundation of Analytics | DBMS | Scalable Databases | ML | Data Mining | Data Algorithms

TECHNICAL SKILLS

Languages: Python, SQL

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly, Statsmodels, XGBoost, TensorFlow (basic)

Big Data & Cloud (SaaS): AWS EMR, EC2, Hive, HDFS, PySpark, GitHub, Google Colab

Modeling & Techniques: Supervised & Unsupervised Learning, Classification & Regression, Time Series Forecasting, Dimensionality Reduction (PCA, SVD), Ensemble Methods, Feature Engineering, Cross-Validation, Hyperparameter Tuning

Data Handling & Tools: BeautifulSoup, Regex, Featuretools, Git, Excel

PROFESSIONAL EXPERIENCE

| | |
|--|----------------------------|
| Data Science Intern – Big Data ML Pipeline | New York, New York |
| Pace University | Mar 2025 – May 2025 |
| <ul style="list-style-type: none">Designed a distributed pipeline on AWS EMR to process 7M+ DOT flight records using Hive and HDFS.Engineered flight features and trained SGD, KNN, and Decision Tree classifiers for delay prediction.Tuned hyperparameters with GridSearchCV, improving F1-score by 26% and reducing false positives by 18%.Identified top delay-heavy airports (ORD, ATL) and airlines (WN), supporting better route planning with data visualization tools. | |
| ML Research Intern - Time Series Forecasting | New York, New York |
| Pace University | April 2025 – May 2025 |
| <ul style="list-style-type: none">Modeled NYC electricity usage trends using Prophet with linear/logistic growth curves on 15 years of data.Forecasted 20-year demand across daily, monthly, and yearly horizons, improving long-term planning accuracy by 38%.Detected seasonal peaks (summer/winter) accounting for 60% of total consumption variation.Enabled a 30% reduction in overcapacity planning errors, supporting smarter grid load allocation. | |
| HeteroGene | Chennai, Tamil Nadu, India |
| Research Intern | Jan 2024 – April 2024 |
| <ul style="list-style-type: none">Developed predictive models for genetic sex identification in Agapornis species, improving accuracy by 25%.Crafted PCR primers and achieved 90% alignment accuracy using NCBI, PubChem, and UniProt datasets.Automated analysis pipelines using Knime and Biovia Studio, reducing processing time by 12% and enhancing reproducibility. | |

SELECTED PROJECTS

| | |
|---|-----------------------|
| NYC Traffic Volume Prediction & Classification | New York, New York |
| Pace University | April 2025 – May 2025 |
| <ul style="list-style-type: none">Built classification and regression models using hourly NYC traffic data to predict congestion levels and total traffic volume.Tuned KNN, SGD Regressor, and Decision Tree models; post-tuning performance improved prediction accuracy by 28%.Uncovered peak-hour traffic contributing to 65% of daily congestion, supporting smarter routing and infrastructure planning. | |
| SQL Logic Programming – UDF | New York, New York |
| Pace University | Feb 2025 – Mar 2025 |
| <ul style="list-style-type: none">Formulated SQL UDFs for automated age-based messaging, mean calculation, and variance analysis by IF-ELSE and WHILE loops.Reduced manual computation time by 40% and improved statistical query accuracy by 22% in simulated business logic pipelines.Enabled real-time data logic execution, scalable across high-volume inputs in analytics workflows. | |