Paridhi Pravin Nigam

pnigam@nyit.edu | +1 551-229-5066 | LinkedIn | GitHub | Location: NY, USA | Open to relocation

Software and data science engineer with 6+ years of experience, proficient in diverse programming languages with 6+ published research papers. Extensive experience working on deployment, ML, and DL models, along with large-scale data processing, feature engineering, and data analytics. Proficient in SQL, Python, Power BI, AWS, R, Alteryx, creating pipelines, cloud technologies, and big data. Currently on EAD work authorization and can work immediately without the need for H1B sponsorship in the future.

EDUCATION

New York Institute of Technology – Manhattan, New York — September 2021 - May 2023
Master of Science in Data Science — GPA: 3.56/4.00
Coursework: Deep Learning | Machine Learning | Data Visualization | Information Retrieval | Optimization Methods | Statistics

North Maharashtra University – Maharashtra, India — August 2016 - June 2018
Master of Engineering in Computer Science and Engineering — GPA: 3.26/4.00
Coursework: Advanced DBMS | Artificial Intelligence| Parallel Computing | Advanced Software Engineering | Distributed System | OOP | Data Structures | Algorithms | Computer Network | NLP | Compiler Design | Operating Systems

***Certifications**:* JPMorgan Chase & Co. Software Engineering Virtual Experience: Software Dev. And Analytics, PwC Power BI Virtual case Experience: Data Analytics and Visualization, CCNA training by Network Bulls.

TECHNICAL SKILLS

Languages and Frameworks: Java, Springboot, Python (pandas, NumPy, seaborn, SciPy, matplotlib, sci-kit-learn, NLTK), R, MATLAB, HTML, Go, CSS, C, C++, JavaScript, JPA, Android, Angular, ReactJS, C#, UML, Cisco Packet Tracer, LINUX, Maven, Spring, Hibernate.
Development environment: Jupyter, OOP, Colab,Visual Studio, Eclipse, IntelliJ IDEA, Anaconda, PyCharm, JIRA, Putty, Flask, AWS, Git, Visual Studio
Visualization Tools: R Studio, Tableau, IBM Cognos Analytics, Power BI, Google charts, Plotly, Datawrapper.
Databases: SQL (Oracle Database, SQLite, MySQL, Firebase), PostgreSQL, NoSQL (MongoDB, Cassandra), Hadoop Map Reduce.
Data warehouses: ClickHouse, Redshift, Salesforce, Databricks, Snowflake, Hudi, Iceberg, Delta Lake, Greenplum.

PROFESSIONAL EXPERIENCE

Data Analyst: Bright Mind Enrichment and Schooling- Remote — August 2023 – Present
· Performed data analysis and maintenance on volunteer databases, ensuring data integrity for websites and their applications for **10k users** across the USA.
· Developed **Tableau dashboards** providing accessible insights for volunteers to optimize fundraising strategies and resource allocation.
· Implemented and streamlined ETL processes for improving overall data quality, **reducing data processing time by 78%.**
· Developed and managed KPI pipelines, delivering actionable insights for strategic decision-making and enhancing fundraising outcomes.

Research Assistant (Data Science): New York Institute of Technology – Manhattan, NY- Onsite — January 2022 – December 2022
· Conducted **data preprocessing** tasks like tokenization, part-of-speech tagging, lemmatization, and NER to determine appropriate data models. Also performed **data scraping** utilizing Beautiful Soup, Tweepy, and Selenium libraries.
· Used **NLU** to understand drone-related data and **NLG** to generate informative and actionable alerts.
· Implemented drone detection/tracking model using binary classification and **long short-term memory (LSTM) - RNN model**, ensuring **93% detection accuracy** and **88% specificity** in safeguarding restricted spaces against multiple potential threats.

Data Analyst Intern: Northeastern University– Boston, MA- Remote — May 2021- June 2021
· Conducted data exploration and analysis on Facebook's open-source data concerning social media-driven online shopping. Utilized R, Excel, and PowerBI to develop interactive reports and predict future sales rates.
· Developed **Tableau visualizations**, including scatter plots, heat maps, bar graphs, and box plots, to monitor data at every stage of its life cycle.
· Applied Random Forest and OpenCV for predictive analysis, **forecasting a 15% sales increase** for 2021 with an **accuracy rate of 92%.**
· Awarded **"The Perfect Planner"** for managing the data analytics team and executing the project within the allocated period with accurate predictions.

Machine Learning Engineer: Deepiotics Pvt.Ltd. – Madhya Pradesh, India May 2019 – August 2021

- Conducted **machine health monitoring analysis** to predict downtimes, identify critical part replacements, plan preventive maintenance, analyze energy consumption, determine spare part quantities, and assess factors influencing production.
- Coordinated data operations utilizing **SQL, SSIS,** and **T-SQL ETLs** to construct an efficient ETL pipeline. Concurrently, delivered live dashboards highlighting current scenarios of production rate, downtime, efficiency, idle time, and rejected items for machines across the factories.
- Applied **predictive and optimization models** to analyze current data from 3 digital factories by leveraging **REST APIs and MLflow**.
- Predicted critical production parameters, identified future production-related factors and risks for management to enable proactive measures.
- Achieved a **40% decrease** in critical failures impacting production rate, **lowered downtime by 35%,** and **boosted throughput by 12%.** Proposed a preventive maintenance schedule and spare part list, resulting in a **$3.3 million increase in sales opportunities and cost savings of $1 million.**

Data Analyst: Government College of Engineering – Maharashtra, India June 2017 – June 2020

- Performed data analysis to detect prevailing trends, patterns, and anomalies for **assessing the profit/loss** associated with expanding a college network.
- Led all phases of data mining, including data collection, cleaning, model development, validation, and visualization for multiple projects.
- **Trained 400+ students** on PHP, Java, data structures, security, OOP, Windows programming, VB.net, and network management.

PROJECT EXPERIENCE

Hospital Readmission Prediction for Heart Failure Patients using ML models

- Conducted **Exploratory Data Analysis** through **visualization and statistical analysis** techniques to extract insights from the heart disease dataset.
- Applied **Random Forest** for **binary classification**, predicting readmission risk for new patients and optimizing model performance through hyperparameter tuning.
- Advocating for adopting automated testing tools to fine-tune model parameters instead of manual methods could **save 90 days of hospital resources, staff time, or patient care.**

5-fold cross validation to evaluate the classification performance of an HMM classifier on Glass Dataset from the UCI machine learning repository

- Successfully implemented a 5-fold cross-validation technique by leveraging **Plotly** and **NumPy** within Python.
- Utilized Python to generate a confusion matrix, yielding a **sensitivity of 79%, specificity of 82%, total accuracy of 82%, an F1-score of 6**, and constructed ROC curve and area under curve (AUC) metrics using the HMM classifier.
- Finalized that the HMM Classifier can be used 5-fold cross-validation for various datasets for classification.

Face Recognition on Labeled Faces in the Wild (LFW) Dataset using Machine Learning and Deep Learning Models

- Enhanced LFW face recognition **accuracy by 75%** through precise model selection, featuring CNNs, Siamese, and Triplet Networks.
- Improved recognition performance on LFW dataset using transfer learning and custom face embeddings, resulting in a **48% reduction in false positives.**

Sentiment Classification of Twitter Data using Supervised Lazy Learning Method

- Determined tweet polarity (positive, negative, and Neutral) using a **lazy learning method, K-nearest neighbors** classifier, and utilized a **Corpus** to handle preprocessed data comprising **1.6 million tweets**.
- Calculated tweet polarity with an **accuracy of 81.1%** and a **specificity of 84.7%.**

Machine learning algorithms on QCM sensor dataset

- Evaluated classification performance for distinguishing between five different types of alcohols using **supervised classifiers** like MLP, SVM, and HMM, alongside **unsupervised classifier** K-means clustering, for each sensor where the HMM classifier achieved the highest accuracy of **90%** among all classifiers.

ACHIEVEMENTS

- Winner in Paper presentation on cloud computing (GCP)
- Participant of Grace Hopper Celebration' 23, Society of Women Engineers 23 and Member of AnitaB.org, Girls Who Code 23
- Salesforce Trailblazer