Jani Miya Shaik

San Diego, CA | janimiyask72@gmail.com | (619)-416-9394
linkedin.com/in/jani-miya-shaik/ | github.com/janimiyarj

PROFESSIONAL EXPERIENCE

LLM & NLP RESEARCH SPECIALIST| *Food and Nutrition Lab, USA*   August 2024 – Present
- Developed an end-to-end data pipeline using **Google Cloud (GCS, Dataflow, BigQuery, Vertex AI**) to securely ingest, preprocess, and transform raw food-related documents, enabling high-quality data preparation for machine learning models.
- Implemented advanced data extraction and quality assessment workflows leveraging Google Document AI's **NLP** and **OCR**, enhancing accuracy in structured data extraction and streamlining feature engineering with Vertex AI's Feature Store.

AI RESEARCH ASSISTANT(LLMs & NLP) | *AIForBusinessLabs, USA*   August 2024 – Present
- Designed advanced **large language models (LLMs)** to analyze gene expression and tissue-specific data, improving biological data interpretation and predictive accuracy.
- Utilized **NLP** techniques, such as **transformers** and **BERT,** to extract data from biomedical literature and genomic datasets, enhancing gene-disease association studies.

DATA SCIENCE INTERN | *Topcon Healthcare, USA*   June 2024 – Present
- Optimized clinical data retrieval with **Python** scripts, achieving a 15% reduction in retrieval time and saving 7 hours weekly through automation, while enhancing scalability with PDF report generation and visualizations.
- Developed machine learning models for predictive analysis and integrated them into **Flask** interfaces, alongside creating an **OpenCV**-based PDF masking tool, which collectively improved efficiency by 12%. Hands-on experience with **PySpark** and exposure to **Azure Databricks** further bolstered data processing capabilities.

GRADUATE TEACHING ASSISTANT | *San Diego State University, USA*   Sep 2023 - Present
- Coordinated 27 students through hands-on data science activities, teaching data cleaning, processing, analysis, and visualization using tools like **Python** and **Tableau,** and demonstrating data storytelling techniques.
- Facilitated **machine learning** projects by assisting students in model training and evaluation on real-world datasets, ensuring statistical rigor with cross-validation and hyperparameter tuning for optimal performance.

DATA PLATFORM SOFTWARE ENGINEER | *Micron Technology***,** *India*   Jul 2021 – Jul 2023
- Led the development of Big Data pipelines using **Python** and **SQL**, achieving a 60% growth, and enabling real-time insights through Business Intelligence dashboards.
- Conducted thorough analysis of scripts to ensure efficient data cleansing, storage, and integration into the Data Warehouse, while setting up **SQL** databases to enhance system performance.
- Innovated a custom business process tool for a **Mobile Application**, reducing manual labor by 13 hours per week and improving overall process efficiency.

DATA SCIENCE INTERN | *Technocolabs, India*   Sep 2020 – Dec 2020
- Managed a collaborative project on intraday trading, applying machine learning techniques and developing an **LSTM** model that achieved **98%** accuracy in stock price forecasting, along with managing **exploratory data analysis** and Flask deployment.

DATA SCIENCE INTERN | *Cognifront, India*   May 2020 – Jun 2020
- Improved loan prediction accuracy by optimizing the **Random Forest** algorithm, achieving an **84%** accuracy rate through precision, recall, and F1 score evaluation, providing actionable insights for decision-making.

TECHNICAL SKILLS

Programming Languages: Python, R, Java, C/C++, JavaScript, HTML5, CSS3, Scala
Databases and Frameworks: SQL Server, PostgreSQL, MongoDB, Flask, React, Angular, Bootstrap, Snowflake, NoSQL
Libraries: TensorFlow, PyTorch, Keras, Matplotlib, Seaborn, PySpark, Plotly, Numpy, Pandas, SciPy, Hive, Databricks
Technologies and Tools: Selenium, AWS, GCP, Tableau, Power BI, ArcGIS, Git, GitHub, Hadoop, Apache Spark
Project Management: Agile methodology, Microsoft Excel, Teamwork, oral communication, presentation, analytical thinking, problem-solving, decision-making, cross-functional, organization

PROJECTS

Glaucoma Detection using Generative Adversarial Networks (GANs):
- Advanced varying dataset resolutions and trained a classification model using the **GAN** network on the Source pictures dataset.
- *Technology Stack:* Data preprocessing, Data modeling, Data visualization, TensorFlow, Recurrent neural network, Data Augmentation

Tourism Analysis:
- Implemented immersive 3D visualizations with **ArcGIS** to identify affordable cities and utilized **Power BI** and **Tableau** for visualizations, highlighting revenue generation and visitor demographics.
- *Technology Stack:* Python, R, Data cleaning, Data organizing, ArcGIS, Tableau, Power BI, Data Manipulation, Data Transformation, Data Acquisition

AI False Image Identification:
- Engineered a robust AI-generated object detector using customized **ResNet50 architecture** and **TensorFlow** achieving 92.2% accuracy in distinguishing real images from algorithmically created ones.
- *Technology Stack*: Exploratory Data Analysis, TensorFlow, Convolutional Neural Network (CNN), Jupyter Notebook

PySpark in Finance: Machine Learning for Loan Approval Prediction:
- Developed a PySpark-based Machine Learning model with 82% AUC to automate loan eligibility prediction, enhancing targeted customer segmentation for informed lending decisions.
- *Technology Stack*: PySpark, MLlib, Seaborn, Matplotlib, Databricks, Scikit-learn, Pandas

Movie Review Sentiment Analysis Engine:
- Engineered a movie sentiment analysis program using **NLTK**, achieving a 63% accuracy.
- *Technology Stack:* Natural Language Processing (NLP), Machine learning, Data Analysis, Text Analysis, Model Training

EDUCATION

Master of Science, Big Data Analytics	Aug 2023 – May 2025
San Diego State University, San Diego, CA	GPA: 3.8/4.0
Bachelor of Engineering, Electronics and Communication	May 2017 – Jul 2021
Chaitanya Bharathi Institute of Technology, India	GPA: 3.8/4.0

Courses Taken: Data Science, Machine Learning, Deep Learning, Mathematics and Statistics, Computer Vision (CV), Artificial Intelligence (AI), Data Structures and Algorithms, Database Management System, Large Language Modelling.