Ganga Surendra Basva

gangasurendrab@gmail.com |+1 856-526-7286 | Maple Shade, NJ | LinkedIn | Github

PROFILE SUMMARY

Data Scientist & Data Engineer with 4+ years of experience designing and deploying data-driven solutions, building scalable ETL pipelines, and implementing machine learning models. Proficient in Python, SQL, Spark, cloud-native platforms (AWS, Azure), and data visualization tools. Experienced in predictive modeling, real-time analytics, and integrating ML pipelines into production systems. Passionate about turning complex datasets into actionable insights.

EDUCATION

Rowan University, Glassboro, NJ August 2023 - May 2025
Master of Science in Data Science | 3.5 GPA

Bharath institute of higher education and research, Chennai, Tamilnadu August 2017 - May 2021
Bachelor of Technology in Computer Science and Engineering | 3.8 GPA

TECHNICAL SKILLS

- **Programming & Analytics:** Python, R, SQL, C#, PySpark, NumPy, Pandas.
- **Machine Learning & AI:** TensorFlow, Keras, PyTorch, Scikit-learn, CNN, Graph Neural Networks.
- **Data Engineering & Cloud:** Spark, Hadoop, Kafka, ETL, AWS (S3, EC2, Lambda, Glue, Kinesis), Azure, Docker, CI/CD.
- **Visualization:** Tableau, Power BI, D3.js, Matplotlib, Seaborn.
- **Databases:** MySQL, PostgreSQL, MongoDB, Redis, Neo4j.
- **Other Tools:** Git, Linux, Windows, VS Code, Agile.

PROFESSIONAL EXPERIENCE

Software Engineer Trainee
Quanti, Philadelphia, USA June 2025 - Present

- Develop an integrated content and ad management platform using Vue.js, FastAPI, and PostgreSQL, enabling seamless posting to multiple social media platforms.
- Develop and optimized RESTful APIs for seamless integration with LinkedIn, X, TikTok, and other social media channels, enabling automated ingestion and processing of content metadata and performance metrics.
- Integrated AI-based content recommendation and optimization modules, leveraging NLP and ML algorithms to enhance user engagement, increasing content interaction metrics by 40%.
- Designed and implemented ETL pipelines for scheduling, processing, and storing high-volume content and ad datasets, ensuring consistency, reliability, and analytics-ready formats.
- Collaborate with product and marketing teams to deliver features on time, using Agile methodologies and CI/CD pipelines for faster releases.

Software Engineer
Tata Elxsi, Bengaluru, India August 2020 - August 2023

- Designed and optimized scalable ETL pipelines and real-time data streaming solutions (Kafka, AWS Kinesis), improving data processing efficiency by 30%.
- Developed and deployed RESTful APIs and microservices for internal data platforms, increasing accessibility and usability of business data by 25%.
- Engineered predictive maintenance solutions using IoT sensor data, resulting in 20% reduction in system downtime.
- Deployed containerized data applications using Docker, ensuring scalable and reproducible analytics environments in production.

Software Developer Intern
Sutherlands, Tamilnadu, India July 2019 - May 2020

- Built automation scripts in Python/Excel macros to streamline customer support processes, reducing handling time by 20%..
- Assisted in building dashboards and visualizations to monitor customer interactions and service efficiency.
- Participated in debugging, feature enhancements, and data validation tasks, gaining hands-on experience with SQL and Python analytics workflows.
- Documented and streamlined data pipelines and reporting processes for internal teams, enhancing accessibility and usability of business intelligence reports.
- Applied data-driven insights to identify bottlenecks and optimize customer support processes, improving overall operational efficiency.

PROJECTS

Early Prediction of Alzheimer's Disease | Python, TensorFlow, Flask API, Keras, CNN January 2019 - April 2019

- Built and trained a CNN model on MRI datasets to detect early-stage Alzheimer's, achieving 85% prediction accuracy.
- Designed a Flask-based REST API for real-time model inference, enabling integration into clinical decision workflows.
- Implemented data preprocessing pipelines including normalization, augmentation, and cleaning to improve model performance.
- Containerized the solution with Docker for scalable deployment and reproducibility in different environments.
- Added robust logging, error handling, and input validation, ensuring production-readiness and reliable predictions in real-world scenarios.

VAST Challenge MC3 | Apache Spark, Hadoop, Graph Neural Networks, Tableau, D3.js September 2023 - November 2023

- Analyzed and processed large-scale datasets (10+ GB structured/unstructured data) using Apache Spark and Hadoop, significantly enhancing data management efficiency.
- Implemented a Graph Neural Network (GNN) service to uncover hidden relationships, boosting insight accuracy by 35%.
- Deployed interactive dashboards with Tableau and D3.js to provide scalable, real-time analytics to end users.
- Optimized data validation, cleansing, and feature engineering pipelines, ensuring high-quality input for analytics and ML models.
- Designed modular, reusable code to support large-scale data processing workflows and accelerate future analytics projects.

NFL Running Back Performance Prediction | Web Scraping, Python, FastAPI, Power BI September 2024 - November 2024

- Collected and analyzed NFL performance datasets (40+ variables) from web sources (NFL, PFF) using Python web scraping techniques, improving dataset quality and completeness.
- Built predictive models (Random Forest, Linear Regression) to forecast player performance and exposed results via FastAPI endpoints.
- Designed an interactive, real-time dashboards using Power BI, resulting in a 25% improvement in strategic decision-making for teams, scouts, and fantasy football users.
- Implemented data validation, error handling, and automated reporting pipelines, ensuring reliable and actionable predictions.
- Optimized backend and ML pipelines for low-latency API responses and scalable processing of large datasets.