Manas Kumar Giri

DATA SCIENTIST

+91- 9178572597 | girimanaskumar1998@gmail.com | GitHub Link | LinkedIn Portfolio

PROFILE SUMMARY

Data Scientist with **2.5 years of experience** in **Machine Learning, NLP, Generative AI, and Data Engineering**. Expertise in **building AI-powered solutions**, predictive analytics, and **MLOps pipelines** across **Azure, AWS, and GCP**. Developed **BERT-based sentiment analysis models**, **Generative AI tools**, and **predictive maintenance systems**, driving efficiency and cost savings. Proficient in **Python, TensorFlow, PyTorch, LangChain, SQL, and Big Data technologies**. Passionate about **solving business challenges** through **data-driven insights**, automation, and scalable AI solutions. **Gold Medalist (MCA) and Kaggle MNIST Top 5% achiever (99.81% accuracy).**

EDUCATION

North Orissa University, Baripada | 2021-23
Master Of Computer Application | **CGPA**: 8.93/10

SKILLS

Programming Languages: Python (Advanced), SQL

Machine Learning: Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, Statistical Analysis, A/B Testing

Deep Learning: Transformers, BERT, GPT, GANs, LSTMs, CNNs, Attention Mechanisms, Autoencoders

NLP & Generative AI: LLM, Hugging Face, LangChain, OpenAI GPT, NLTK, SpaCy, RAG, Agentic AI, Vertex AI

Data Engineering: ETL Pipelines, Feature Engineering, Spark, Kafka, Snowflake

Cloud Platforms: Azure (ML Studio, Data Factory, Synapse Analytics, Databricks), AWS (SageMaker, Lambda, S3, EC2)

MLOps: Docker, Kubernetes, MLflow, Git, GitHub Actions, CI/CD Pipelines, DVC (Data Version Control)

Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly, Alteryx

Databases & Storage: MySQL, Cassandra, NoSQL

Soft Skills: Analytical Thinking, Problem-Solving, Team Collaboration, Stakeholder Communication

WORK EXPERIENCE

DATA SCIENTIST

REGIS SOFT Pvt. Ltd., Pune | June 2023 – May 2025

- **AI and NLP Solutions:** Developed a **BERT-based sentiment analysis model** using Azure Cognitive Services, improving accuracy by **25%**, leading to enhanced customer satisfaction strategies. Built an **NLP pipeline with BERT & SpaCy**, achieving **92% accuracy** and improving product satisfaction by **15%.**

- **Generative AI Applications:** Created a content generation tool using **OpenAI GPT & LangChain**, reducing content creation time by **40%** and increasing engagement by **20%**.

- **Predictive Maintenance:** Designed and deployed a **predictive maintenance model**, reducing downtime by **30%** and saving operational costs.

- **Machine Learning & Data Science:** Developed a **customer segmentation model** using **K-Means & DBSCAN**, optimizing marketing campaigns and improving efficiency by **25%**. Conducted **A/B testing** for product features, providing data-driven insights to enhance user engagement.

- **ML Pipeline Optimization:** Built scalable **ML pipelines** in **Azure ML Studio & Databricks,** reducing lead time by **40%.**

- **Cloud Integration & Deployment:** Led the migration of legacy analytics systems to **Azure Synapse Analytics & Data Factory**, enabling real-time data processing. Deployed ML models on **AWS SageMaker**, reducing inference time by **30%**.

- **ETL & Big Data Processing:** Automated ETL workflows, **reducing data pipeline** latency by **60%** and improving efficiency.

- **Data Visualization:** Created **interactive dashboards** in **Power BI & Tableau** for KPI tracking, enabling data-driven decisions and contributing to a **10% revenue increase**.

- **Collaboration:** Partnered with cross-functional teams to integrate machine learning & AI solutions.

- **Key Technologies:** Python, Azure ML, Databricks, PyTorch, TensorFlow, BERT, OpenAI GPT, LangChain, Power BI, Tableau, Azure Synapse Analytics, Azure Data Factory, AWS SageMaker, SQL, NLP, Machine Learning, ETL Automation, Big Data Processing

DATA SCIENCE INTERN

REGIS SOFT Pvt. Ltd., Pune | December 2022 – May 2023

- **Exploratory Data Analysis (EDA):** Conducted in-depth EDA on large datasets, improving model performance by **30%** through informed feature engineering.
- **Recommendation Engine:** Built and deployed a **recommendation engine** for e-commerce platforms, enhancing user retention by **20%**.
- **Sentiment Analysis:** Collaborated on a Natural Language Processing (NLP) project, achieving **92% accuracy** in analyzing customer feedback.
- **Anomaly Detection:** Worked on anomaly detection models, identifying **98% of fraudulent transactions** in financial data.
- **Interactive Dashboards:** Created **interactive dashboards** using Tableau and Matplotlib, effectively communicating analytical results to stakeholders.

- **Key Technologies:** Python, NLP, Tableau, TensorFlow, PyTorch, Scikit-learn

PROJECTS

Project 1: AI-Powered Sentiment Analysis System

Client: Aretto (Emerging Indian e-commerce brand)

Description:
Developed a real-time sentiment analysis model using **BERT** and **Azure Cognitive Services**, improving accuracy by **25%** and boosting product satisfaction by **15%**.

Responsibilities:

- Fine-tuned BERT with domain data via Hugging Face & Azure ML.

- Built NLP pipeline with SpaCy and deployed via FastAPI.

- Visualized results using Power BI for marketing insights.

Tools: Python, BERT, Hugging Face, Azure ML Studio, SpaCy, FastAPI, Power BI

Project 2: Generative AI Content Automation Tool

Client: Murf AI (AI, Voice Technology)

Description:
Built a content generation tool using **OpenAI GPT** and **LangChain**, reducing manual effort by **40%** and increasing engagement by **20%**.

Responsibilities:

- Created prompt workflows and integrated LangChain.

- Deployed FastAPI-based API linked to CMS.

- Conducted A/B testing for performance tracking.

Tools: Python, OpenAI GPT, LangChain, FastAPI, Prompt Engineering, A/B Testing, Power BI

Project 3: Fraud Detection System for Financial Transactions

Client: Rupifi (Fintech, B2B Payments)

Description:
Built a real-time anomaly detection system for fraudulent transactions, achieving **98%** accuracy and reducing financial losses by **35%**.

Responsibilities:

- Developed anomaly detection models using Isolation Forest and XGBoost.

- Engineered features with Pandas and PySpark, deployed via Azure ML and FastAPI.

- Visualized fraud alerts with Power BI and automated pipelines with Azure Data Factory.

Tools: Python, Isolation Forest, XGBoost, Pandas, PySpark, Azure Machine Learning, FastAPI, MLflow, Azure Data Factory, Power BI

Project 4: Predictive Analytics and Generative AI for Healthcare Billing

Client: Apollo HealthTech (Healthcare, Digital Health Solutions)

Description:
Created an ML and RAG-based system for patient billing and fraud detection, cutting fraudulent claims by 28% and speeding up billing by 40%.

Responsibilities:

- Built predictive models with RandomForest, LSTM and RAG pipeline with LangChain.

- Deployed via AWS SageMaker with FastAPI, using Pinecone for document retrieval.

- Automated ETL with AWS Glue and visualized insights in Tableau.

Tools: Python, Random Forest, LSTM, LangChain, RAG, Hugging Face, Transformers, Pinecone, AWS SageMaker, FastAPI, AWS Glue, Databricks, Tableau

Project 5: Scalable ML Pipeline and Deployment System

Client: Gong Labs (Sales Tech, Revenue Intelligence)

Description:
Designed and deployed scalable ML pipelines on **AWS & Azure**, reducing inference time by **30%** and ETL latency by **60%**.

Responsibilities:

- Built automated pipelines in Azure ML & Databricks.

- Deployed via SageMaker and monitored using MLflow.

- Enabled CI/CD with GitHub Actions.

Tools: Python, Databricks, Azure ML, AWS SageMaker, MLflow, Azure Data Factory, GitHub Actions, PySpark

CERTIFICATIONS & ACHIEVEMENTS

- **Full Stack Data Science** - iNeuron
- **Alteryx Machine Learning Fundamentals Certified** - Alteryx
- **AI/ML Development** - GUVI
- **Pro5** - Ranked in the top **5%** in Pro5's Machine Learning assessment, showcasing expertise in ML algorithms.
- **Publications:** Published a blog on "Generative AI, ML & DL" on **Medium**.
- **Kaggle MNIST Competition** - Achieved **99.81%** accuracy (Top 5% globally).