

## SHARMI DAS

Mobile: +14376625715 | E-Mail: dassharmi6@gmail.com

LinkedIn: www.linkedin.com/in/sharmidas0402 | GitHub: github.com/Sharmidas201

Address: Toronto, ON, Canada

## SUMMARY

Results-driven Data Scientist with a completed MSc in Data Science and Analytics and experience building scalable data pipelines, predictive models, and production-level machine learning deployments. Skilled in Python, PySpark, SQL, and cloud platforms (AWS, GCP, Azure). Adept at transforming structured and unstructured data into actionable insights through automation, reporting, and visualization. Dedicated to applying AI and analytics to solve real-world business challenges across healthcare, finance, and operations.

## EDUCATION

### **-MSc in Data Science and Analytics, Toronto Metropolitan University(Completed-2025)**

Vector Scholarship in AI | Hack the World Hackathon Winner | CGPA: 4.28/4.33

### **-Ontario Graduate Certificate in Artificial Intelligence and Machine Learning, Humber College (2024)**

Dean's Honour List | Grade: 92.5%

### **-Bachelor of Technology in Computer Science and Engineering, Future Institute of Technology, Kolkata (2023)**

CGPA: 9.2/10

## TECHNICAL SKILLS

- **Programming & Data Tools:** Python (NumPy, Pandas, Scikit-learn), PySpark, SQL (HiveQL, Spark SQL), Java, C++, Git, Selenium, HTML/CSS/JavaScript, LangChain
- **Machine Learning & AI:** TensorFlow, PyTorch, XGBoost, Random Forest, CNNs, NLP/NLU (BERT, spaCy), Generative AI, LLMs, Document Classification, Entity Recognition, Model Training & Deployment, Prompt Engineering
- **Data Engineering & Cloud:** AWS (EC2, S3, Lambda, SageMaker), GCP, Azure (Notebooks, OpenAI), Databricks, Hadoop, Hive, ETL Pipelines, Distributed Computing
- **Visualization & Reporting:** Tableau, Kibana, Matplotlib, Seaborn, Data Storytelling, Technical Documentation
- **Soft Skills:** Communication, Adaptability, Team Collaboration, Documentation, Creative Problem-Solving

## WORK EXPERIENCE

### **Graduate Assistant | Toronto Metropolitan University | Winter 2025&Summer 2025**

1. Supported instruction in six graduate/undergraduate courses by supervising labs, grading, and preparing technical materials.
2. Provided real-time technical support to 200+ students, ensuring smooth lab operations and resolving assignment/experiment challenges.

### **Cognizant - Quality Engineering & Assurance Intern**

February,2023-July,2023

1. Automated car website testing using Selenium-Java, cutting manual QA effort by 30% and improving site performance by 25%.
2. Collaborated with a cross-functional team of 6 engineers to streamline test coverage and improve defect detection.

## MACHINE LEARNING AND DEEP LEARNING PROJECTS

### Healthcare AI & Predictive Modeling

- Developed an LSTM model for epilepsy seizure prediction with 95% accuracy, securing \$5100 funding for its potential to enhance patient care and response times.
- Deployed a machine learning pipeline on AWS for stroke prediction using XGBoost, achieving 94.7% accuracy and enabling scalable clinical risk assessment.[\[Github\]](#)
- Trained a CNN with PyTorch/ResNet to classify Alzheimer's stages from MRI images, reaching 86% accuracy and supporting early detection efforts.[\[Github\]](#)

### Big Data Analytics & Data Engineering

- Designed an end-to-end Hadoop–PySpark–Hive–Kibana pipeline analyzing 22,000+ energy records across 129 features, uncovering GDP-linked transition gaps (high-GDP countries consume 73% more non-renewable energy) to inform sustainability policies.[\[Github\]](#)
- Created regression and RNN-based forecasting models for profits and sales, delivering high predictive performance ( $R^2 = 0.90$ , MAE = 0.197).[\[Github\]](#)

### Generative AI & NLP Systems

- Built a Retrieval-Augmented Generation (RAG) chatbot with LangChain, Pinecone, and Llama 2 integration, improving domain-specific scientific QA by enhancing retrieval accuracy and reducing irrelevant responses.[\[Github\]](#)
- Reduced jewelry design iteration time by ~40% by developing Generative AI models for rapid prototyping of unique ring styles.[\[Github\]](#)

### Applied Research – Major Research Paper (MRP)

- Analyzed 7,900+ TCGA patient records, integrating mutation, expression, and survival data to identify entropy-based biomarkers for cancer prognosis.[\[Github\]](#)
- Built Cox models with entropy-derived pathway metrics, uncovering key prognostic pathways (e.g., MTOR→RB) with statistically significant survival associations
- Demonstrated entropy's potential as an interpretable biomarker for adaptive therapy design, providing a foundation for future deep learning integration in cancer prognosis.